CENGAGE Learning™

Probabilidad y estadística

para ingeniería y ciencias

Octava edición

Jay L. Devore

Probabilidad y estadística para ingeniería y ciencias

OCTAVA EDICIÓN

Probabilidad y estadística para ingeniería y ciencias

JAY DEVORE

California Polytechnic State University, San Luis Obispo

Traducción:
Patricia Solorio Gómez
Traductora profesional

Revisión técnica:
Ana Elizabeth García Hernández
Universidad LaSalle Morelia

CENGAGE Learning™

Australia • Brasil • Corea • España • Estados Unidos • Japón • México • Reino Unido • Singapur

CENGAGE Learning™

Probabilidad y estadística para ingeniería y ciencias, Octava edición
Jay L. Devore

Presidente de Cengage Learning Latinoamérica:
Fernando Valenzuela Migoya

Director de producto y desarrollo Latinoamérica:
Daniel Oti Yvonnet

Director editorial y de producción Latinoamérica:
Raúl D. Zendejas Espejel

Editor:
Sergio R. Cervantes González

Coordinadora de producción editorial:
Abril Vega Orozco

Editor de producción:
Timoteo Eliosa García

Coordinador de manufactura:
Rafael Pérez González

Diseño de portada:
Rokusek Design

Composición tipográfica:
Imagen Editorial

Impreso en México
1 2 3 4 5 6 7 15 14 13 12

Traducido del libro:
Probability and Statistics for Engineering and Sciences, Eighth Edition
Jay L. Devore
Publicado en inglés por Brooks/Cole, Cengage Learning, © 2010
ISBN-13: 978-0-538-73352-6
ISBN-10: 0-538-73352-7

Datos para catalogación bibliográfica:
Devore, Jay L.
Probabilidad y estadística para ingeniería y ciencias, Octava edición
ISBN: 978-607-481-619-8

Visite nuestro sitio en:
http://latinoamerica.cengage.com

Para mi nieto
Philip, quien es
estadísticamente
significativo.

Contenido

11 Análisis multifactorial de la varianza

12 Regresión lineal simple y correlación

13 Regresión múltiple y no lineal

14 Pruebas de bondad de ajuste y análisis de datos categóricos

Prefacio

Propósito

El uso de modelos de probabilidad y métodos estadísticos para analizar datos se ha convertido en una práctica común en virtualmente todas las disciplinas científicas. Este libro pretende introducir con amplitud aquellos modelos y métodos que con mayor probabilidad se encuentran y utilizan los estudiantes en sus carreras de ingeniería y las ciencias naturales. Aun cuando los ejemplos y ejercicios se diseñaron pensando en los científicos e ingenieros, la mayoría de los métodos tratados son básicos en los análisis estadísticos en muchas otras disciplinas, por lo que los estudiantes de las ciencias administrativas y sociales también se beneficiarán con la lectura del libro.

Enfoque

Los estudiantes de un curso de estadística diseñado para servir a otras especialidades de estudio al principio es posible que duden del valor y relevancia del material, pero mi experiencia es que los estudiantes *pueden* ser conectados a la estadística con el uso de buenos ejemplos y ejercicios que combinen sus experiencias diarias con sus intereses científicos. Así pues, he trabajado duro para encontrar ejemplos reales y no artificiales, que alguien pensó que valía la pena recopilar y analizar. Muchos de los métodos presentados, sobre todo en los últimos capítulos sobre inferencia estadística, se ilustran analizando datos tomados de una fuente publicada y muchos de los ejercicios también implican trabajar con dichos datos. En ocasiones es posible que el lector no esté familiarizado con el contexto de un problema particular (como muchas veces yo lo estuve), pero me di cuenta que los problemas reales con un contexto un tanto extraño atraen más a los estudiantes que aquellos problemas definitivamente artificiales en un entorno conocido.

Nivel matemático

La exposición es relativamente modesta en función de desarrollo matemático. El uso sustancial del cálculo se hace sólo en el capítulo 4 y en partes de los capítulos 5 y 6. En particular, con excepción de una observación o nota ocasional, el cálculo aparece en la parte de inferencia del libro sólo en la segunda sección del capítulo 6. No se utiliza álgebra matricial en absoluto. Por lo tanto casi toda la exposición deberá de ser accesible para aquellos cuyo conocimiento matemático incluye un semestre o dos trimestres de cálculo diferencial e integral.

Contenido

El capítulo 1 se inicia con algunos conceptos y terminología básicos (población, muestra, estadística descriptiva e inferencial, estudios enumerativos contra analíticos, y así sucesivamente) y continúa con el estudio de métodos descriptivos gráficos y numéricos importantes. En el capítulo 2 se da un desarrollo un tanto tradicional de la probabilidad, seguido por distribuciones de probabilidad de variables aleatorias continuas y discretas en los capítulos 3 y 4, respectivamente. Las distribuciones conjuntas y sus propiedades se analizan en la primera parte del capítulo 5. La última parte de este capítulo introduce la estadística y sus distribuciones muestrales, las cuales constituyen el puente entre probabilidad e inferencia. Los siguientes tres capítulos se ocupan de la estimación puntual, los intervalos estadísticos y la comprobación de hipótesis basados en una muestra única. Los métodos de inferencia que implican dos muestras independientes y datos apareados se presentan en el capítulo 9. El análisis de la varianza es el tema de los capítulos 10 y 11 (unifactorial y multifactorial, respectivamente). La regresión aparece por primera vez en el capítulo 12 (el

modelo de regresión lineal simple y correlación) y regresa para una amplia repetición en el capítulo 13. Los últimos tres capítulos analizan métodos ji al cuadrado, procedimientos sin distribución (no paramétricos) y técnicas de control estadístico de calidad.

Ayuda para el aprendizaje de los estudiantes

Aunque el nivel matemático del libro representará poca dificultad para la mayoría de los estudiantes de ciencias e ingeniería, es posible que el trabajo dirigido hacia la comprensión de los conceptos y apreciación del desarrollo lógico de la metodología en ocasiones requiera un esfuerzo sustancial. Para ayudar a que los estudiantes ganen en comprensión y apreciación he proporcionado numerosos ejercicios de dificultad variable, desde muchos que implican la aplicación rutinaria del material incluido en el texto hasta algunos que piden al lector que extienda los conceptos analizados en el texto a situaciones un tanto nuevas. Existen muchos ejercicios que la mayoría de los profesores desearían asignar durante cualquier curso particular, pero recomiendo que se les pida a los estudiantes que resuelvan un número sustancial de ellos; en una disciplina de solución de problemas, el compromiso activo de esta clase es la forma más segura de identificar y cerrar las brechas en el entendimiento que inevitablemente surgen. Las respuestas a la mayoría de los ejercicios impares aparecen en la sección de respuestas al final del texto.

Para acceder a material adicional del curso y recursos de apoyo, por favor visite www.cengagebrain.com. Éste le llevará a la página en donde encontrará material de apoyo para el libro.

Nuevo en esta edición

- Un glosario de los símbolos y abreviaturas aparece al final del libro (el autor se disculpa por su pereza de no proporcionar este conjunto en ediciones anteriores) y un pequeño conjunto de exámenes de muestra aparece en la página web del libro (disponible en www.cengage.com/login).
- Muchos ejemplos nuevos y ejercicios, casi todos basados en datos reales o problemas reales. Algunos de estos escenarios son menos técnicos o de alcance más amplio que lo que ha sido incluido en las ediciones anteriores; por ejemplo, el peso de los jugadores de futbol (para ilustrar la multimodalidad), los gastos de recaudación de fondos para organizaciones caritativas y la comparación de los promedios de calificaciones en las clases impartidas por profesores a tiempo parcial con los de las clases impartidas por profesores de tiempo completo.
- El material de los valores de P, ha sido reescrito. El valor P es ahora definido inicialmente como una probabilidad más que como el menor nivel de importancia para los que puede ser rechazada la hipótesis nula. Se presenta un experimento de simulación para ilustrar el comportamiento de los valores de P.
- El capítulo 1 contiene una nueva subsección sobre “El alcance de la estadística moderna” para indicar cómo los profesionales de la estadística siguen desarrollando nueva metodología de trabajo, mientras trabajan en problemas en un amplio espectro de disciplinas.
- La exposición ha sido pulida siempre que sea posible para ayudar a los estudiantes a adquirir una comprensión intuitiva de los diferentes conceptos. Por ejemplo, la función de distribución acumulada es presentada deliberadamente en el capítulo 3, el primer ejemplo de máxima verosimilitud en la sección 6.2 contiene una discusión más cuidadosa de la probabilidad, se presta más atención a la potencia y probabilidades de error tipo II en la sección 8.3, y el material de residuos y las sumas de cuadrados de regresión múltiple se presenta de manera más explícita en la sección 13.4.

Reconocimentos

Mis colegas en Cal Poly me proporcionaron apoyo y retroalimentación invaluables durante el curso de los años. También agradezco a los muchos usuarios de ediciones previas que me sugirieron mejoras (y en ocasiones errores identificados). Una nota especial de agradecimiento va para Matt Carlton por su trabajo en los dos manuales de soluciones, uno para profesores y el otro para estudiantes.

La generosa retroalimentación provista por los siguientes revisores de esta y ediciones previas, ha sido de mucha ayuda para mejorar el libro: Robert L. Armacost, University of Central Florida; Bill Bade, Lincoln Land Community College; Douglas M. Bates, University of Wisconsin–Madison; Michael Berry, West Virginia Wesleyan College; Brian Bowman, Auburn University; Linda Boyle, University of Iowa; Ralph Bravaco, Stonehill College; Linfield C. Brown, Tufts University; Karen M. Bursic, University of Pittsburgh; Lynne Butler, Haverford College; Raj S. Chhikara, University of Houston–Clear Lake; Edwin Chong, Colorado State University; David Clark, California State Polytechnic University at Pomona; Ken Constantine, Taylor University; David M. Cresap, University of Portland; Savas Dayanik, Princeton University; Don E. Deal, University of Houston; Annjanette M. Dodd, Humboldt State University; Jimmy Doi, California Polytechnic State University–San Luis Obispo; Charles E. Donaghey, University of Houston; Patrick J. Driscoll, U.S. Military Academy; Mark Duva, University of Virginia; Nassir Eltinay, Lincoln Land Community College; Thomas English, College of the Mainland; Nasser S. Fard, Northeastern University; Ronald Fricker, Naval Postgraduate School; Steven T. Garren, James Madison University; Mark Gebert, University of Kentucky; Harland Glaz, University of Maryland; Ken Grace, Anoka-Ramsey Community College; Celso Grebogi, University of Maryland; Veronica Webster Griffis, Michigan Technological University; Jose Guardiola, Texas A&M University–Corpus Christi; K. L. D. Gunawardena, University of Wisconsin–Oshkosh; James J. Halavin, Rochester Institute of Technology; James Hartman, Marymount University; Tyler Haynes, Saginaw Valley State University; Jennifer Hoeting, Colorado State University; Wei-Min Huang, Lehigh University; Aridaman Jain, New Jersey Institute of Technology; Roger W. Johnson, South Dakota School of Mines & Technology; Chihwa Kao, Syracuse University; Saleem A. Kassam, University of Pennsylvania; Mohammad T. Khasawneh, State University of NewYork–Binghamton; Stephen Kokoska, Colgate University; Hillel J. Kumin, University of Oklahoma; Sarah Lam, Binghamton University; M. Louise Lawson, Kennesaw State University; Jialiang Li, University of Wisconsin–Madison; Wooi K. Lim, William Paterson University; Aquila Lipscomb, The Citadel; Manuel Lladser, University of Colorado at Boulder; Graham Lord, University of California–Los Angeles; Joseph L. Macaluso, DeSales University; Ranjan Maitra, Iowa State University; David Mathiason, Rochester Institute of Technology; Arnold R. Miller, University of Denver; John J. Millson, University of Maryland; Pamela Kay Miltenberger, West Virginia Wesleyan College; Monica Molsee, Portland State University; Thomas Moore, Naval Postgraduate School; Robert M. Norton, College of Charleston; Steven Pilnick, Naval Postgraduate School; Robi Polikar, Rowan University; Ernest Pyle, Houston Baptist University; Steve Rein, California Polytechnic State University–San Luis Obispo; Tony Richardson, University of Evansville; Don Ridgeway, North Carolina State University; Larry J. Ringer, Texas A&M University; Robert M. Schumacher, Cedarville University; Ron Schwartz, Florida Atlantic University; Kevan Shafizadeh, California State University–Sacramento; Mohammed Shayib, Prairie View A&M; Robert K. Smidt, California Polytechnic State University–San Luis Obispo; Alice E. Smith, Auburn University; James MacGregor Smith, University of Massachusetts; Paul J. Smith, University of Maryland; Richard M. Soland, The George Washington University; Clifford Spiegelman, Texas A&M University; Jery Stedinger, Cornell University; David Steinberg, Tel Aviv University;

William Thistleton, State University of New York Institute of Technology; G. Geoffrey Vining, University of Florida; Bhutan Wadhwa, Cleveland State University; Gary Wasserman, Wayne State University; Elaine Wenderholm, State University of New York–Oswego; Samuel P. Wilcock, Messiah College; Michael G. Zabetakis, University of Pittsburgh, y Maria Zack, Point Loma Nazarene University.

Danielle Urban de Elm Street Publishing Services ha realizado un trabajo excelente al supervisar la producción del libro. Una vez más me veo obligado a expresar mi gratitud a todas aquellas personas en Cengage que han hecho contribuciones importantes a lo largo de mi carrera como escritor de libros de texto. Para esta edición más reciente, un agradecimiento especial a Jay Campbell (por su información oportuna y retroalimentación a través del proyecto), Molly Taylor, Shaylin Walsh, Ashley Pickering, Cathy Brooks y Andrew Coppola. También apreciamos la labor estelar de todos los representantes de ventas de Cengage Learning que han trabajado para hacer que mis libros sean más visibles para la comunidad estadística. Por último, pero no por ello menor, un sincero agradecimiento a mi esposa Carol por sus décadas de apoyo, y a mis hijas por proporcionar inspiración a través de sus propios logros.

Jay Devore

1 Generalidades y estadística descriptiva

"No soy muy dado a lamentar, pero estuve desconcertado sobre esto un tiempo. Creo que debería haber estudiado mucho más estadísticas en la universidad."

Max Levchin, cofundador de Paypal, fundador de Slide.

Cita de la semana tomada del sitio web de la
American Statistical Association, 23 de noviembre de 2010

"Sigo diciendo que el trabajo sexy en los próximos 10 años serán los estadísticos y no estoy bromeando."

Hal Varian, economista en jefe Google.

The New York Times, 6 de agosto de 2009.

INTRODUCCIÓN

Los conceptos y métodos estadísticos no son sólo útiles sino que con frecuencia son indispensables para entender el mundo que nos rodea. Proporcionan formas de obtener ideas nuevas del comportamiento de muchos fenómenos que se presentarán en su campo de especialización escogido en ingeniería o ciencia.

La disciplina de estadística nos enseña cómo realizar juicios inteligentes y tomar decisiones informadas en la presencia de incertidumbre y variación. Sin incertidumbre o variación, habría poca necesidad de métodos estadísticos o de profesionales en estadística. Si cada componente de un tipo particular tuviera exactamente la misma duración, si todos los resistores producidos por un fabricante tuvieran el mismo valor de resistencia, si las determinaciones del pH en muestras de suelo de un lugar particular dieran resultados idénticos, y así sucesivamente, entonces una sola observación revelaría toda la información deseada.

Una importante manifestación de variación surge en el curso de la medición de emisiones en vehículos automotores. Los requerimientos de costo y tiempo del Federal Test Procedure (FTP) impiden su uso generalizado en programas de inspección de vehículos. En consecuencia, muchas agencias han creado pruebas menos costosas y más rápidas, las que se espera reproduzcan los resultados obtenidos con el FTP. De acuerdo con el artículo "Motor Vehicle Emissions Variability" (*J. of the Air and*

Waste Mgmt. Assoc., 1996: 667-675), la aceptación del FTP como patrón de oro ha llevado a la creencia ampliamente difundida de que las mediciones repetidas en el mismo vehículo conducirían a resultados idénticos (o casi idénticos). Los autores del artículo aplicaron el FTP a siete vehículos caracterizados como "altos emisores". He aquí los resultados de uno de los vehículos.

HC (g/milla)	13.8	18.3	32.2	32.5
CO (g/milla)	118	149	232	236

La variación sustancial en las mediciones tanto de HC como de CO proyecta una duda considerable sobre la sabiduría convencional y hace mucho más difícil realizar evaluaciones precisas sobre niveles de emisiones.

¿Cómo se pueden utilizar técnicas estadísticas para reunir información y sacar conclusiones? Supóngase, por ejemplo, que un ingeniero de materiales inventó un recubrimiento para retardar la corrosión en tuberías de metal en circunstancias específicas. Si este recubrimiento se aplica a diferentes segmentos de la tubería, la variación de las condiciones ambientales y de los segmentos mismos producirá más corrosión sustancial en algunos segmentos que en otros. Se podría utilizar un análisis estadístico en datos de dicho experimento para decidir si la cantidad *promedio* de corrosión excede un límite superior especificado de alguna clase o para predecir cuánta corrosión ocurrirá en una sola pieza de tubería.

Por otra parte, supóngase que el ingeniero inventó el recubrimiento con la creencia de que será superior al recubrimiento actualmente utilizado. Se podría realizar un experimento comparativo para investigar esta cuestión aplicando el recubrimiento actual a algunos segmentos de la tubería y el nuevo a otros segmentos. Esto debe realizarse con cuidado o se obtendrá una conclusión errónea. Por ejemplo, tal vez la cantidad promedio de corrosión sea idéntica con los dos recubrimientos. Sin embargo, el recubrimiento nuevo puede ser aplicado a segmentos que tengan una resistencia superior a la corrosión y en condiciones ambientales menos severas en comparación con los segmentos y condiciones del recubrimiento actual. El investigador probablemente observaría entonces una diferencia entre los dos recubrimientos atribuibles no a los recubrimientos mismos, sino sólo a variaciones extrañas. La estadística ofrece no sólo métodos para analizar resultados de experimentos una vez que se han realizado sino también sugerencias sobre cómo pueden realizarse los experimentos de una manera eficiente para mitigar los efectos de la variación y tener una mejor oportunidad de llegar a conclusiones correctas.

1.1 Poblaciones, muestras y procesos

Los ingenieros y científicos constantemente están expuestos a la recolección de hechos o **datos**, tanto en sus actividades profesionales como en sus actividades diarias. La disciplina de la estadística proporciona métodos de organizar y resumir datos y de sacar conclusiones basadas en la información contenida en los datos.

Una investigación típicamente se enfocará en una colección bien definida de objetos que constituyen una **población** de interés. En un estudio, la población podría consistir de todas las cápsulas de gelatina de un tipo particular producidas durante un periodo específico. Otra investigación podría implicar la población compuesta de todos los individuos que recibieron una licenciatura de ingeniería durante el año académico más reciente. Cuando la información deseada está disponible para todos los objetos de la población, se tiene lo que se llama un **censo**. Las restricciones de tiempo, dinero y otros recursos escasos casi siempre hacen que un censo sea impráctico o infactible. En su lugar, se selecciona un subconjunto de la población, **una muestra**, de manera pre-escrita. Así pues, se podría obtener una muestra de cojinetes de una corrida de producción particular como base para investigar si los cojinetes se ajustan a las especificaciones de fabricación, o se podría seleccionar una muestra de los graduados de ingeniería del último año para obtener retroalimentación sobre la calidad de los programas de estudio de ingeniería.

Por lo general existe interés sólo en ciertas características de los objetos en una población: el número de grietas en la superficie de cada recubrimiento, el espesor de cada pared de cápsula, el género de un graduado de ingeniería, la edad a la cual el individuo se graduó, y así sucesivamente. Una característica puede ser categórica, tal como el género o tipo de funcionamiento defectuoso o puede ser de naturaleza numérica. En el primer caso, el *valor* de la característica es una categoría (p. ej., femenino o soldadura insuficiente), mientras que en el segundo caso, el valor es un número (p. ej., edad = 23 años o diámetro = .502 cm). Una **variable** es cualquier característica cuyo valor puede cambiar de un objeto a otro en la población. Inicialmente las letras minúsculas del final de nuestro alfabeto denotarán las variables. Algunos ejemplos incluyen:

x = marca de la calculadora de un estudiante

y = número de visitas a un sitio web particular durante un periodo específico

z = distancia de frenado de un automóvil en condiciones específicas

Se obtienen datos al observar o una sola variable o en forma simultánea dos o más variables. Un conjunto de datos **univariantes** se compone de observaciones realizadas en una sola variable. Por ejemplo, se podría determinar el tipo de transmisión automática (A) o manual (M) en cada uno de diez automóviles recientemente adquiridos en cierto concesionario y el resultado sería el siguiente conjunto de datos categóricos

M A A A M A A M A A

La siguiente muestra de duraciones (horas) de baterías de la marca D puestas en cierto uso es un conjunto de datos numéricos univariantes:

5.6 5.1 6.2 6.0 5.8 6.5 5.8 5.5

Se tienen datos **bivariantes** cuando se realizan observaciones en cada una de dos variables. El conjunto de datos podría consistir en un par (altura, peso) por cada jugador integrante del equipo de basquetbol, con la primera observación como (72, 168), la segunda como (75, 212), y así sucesivamente. Si un ingeniero determina el valor tanto de x = componente de duración y y = razón de la falla del componente, el conjunto de datos resultante es bivariante con una variable numérica y la otra categórica. Los datos **multivariantes** surgen cuando se realizan observaciones en más de una variable (por lo que bivariante es un caso especial de multivariante). Por ejemplo, un médico investigador podría determinar la presión sanguínea sistólica, la presión sanguínea diastólica y el nivel de colesterol en suero de cada paciente participante en un estudio. Cada observación sería una terna de números, tal como (120, 80, 146). En muchos conjuntos de datos multivariantes, algunas variables son numéricas y otras son categóricas. Por lo tanto el número anual dedicado al automóvil de *Consumer Reports* da valores de tales variables como tipo de vehículo (pequeño, deportivo, compacto, tamaño mediano, grande), eficiencia de consumo de combustible en la ciudad y en carretera en millas por galón (mpg), tipo de tren motriz (ruedas traseras, ruedas delanteras, cuatro ruedas), etcétera.

Ramas de la estadística

Es posible que un investigador que ha recopilado datos desee resumir y describir características importantes de los mismos. Esto implica utilizar métodos de **estadística descriptiva**. Algunos de ellos son de naturaleza gráfica; la construcción de histogramas, diagramas de caja y gráficas de puntos son ejemplos primordiales. Otros métodos descriptivos implican el cálculo de medidas numéricas, tales como medias, desviaciones estándar y coeficientes de correlación. La amplia disponibilidad de programas de computadora estadísticos han hecho que estas tareas sean más fáciles de realizar de lo que antes eran. Las computadoras son mucho más eficientes que los seres humanos para calcular y crear imágenes (¡una vez que han recibido las instrucciones apropiadas del usuario!). Esto significa que el investigador no tiene que esforzarse mucho en el "trabajo tedioso" y tendrá más tiempo para estudiar los datos y extraer mensajes importantes. A lo largo de este libro se presentarán los datos de salida de varios paquetes tales como Minitab, SAS, S-Plus y R. El programa R puede ser descargado sin cargo del sitio http://www.r-project.org.

Ejemplo 1.1 La caridad es un gran negocio en Estados Unidos. El sitio web charitynavigator.com proporciona información de aproximadamente 5500 organizaciones de caridad y existe un gran número de pequeñas caridades que vuelan debajo del radar de la pantalla del navegador. Algunas organizaciones caritativas operan de modo muy eficiente, con gastos administrativos y de recaudación de fondos que sólo son un pequeño porcentaje de los gastos totales, mientras que otras gastan un alto porcentaje de lo que pueden tomar en tales actividades. En seguida se muestran los datos de los gastos en la recaudación de fondos como un porcentaje de los gastos totales para una muestra aleatoria de 60 organizaciones de caridad:

6.1	12.6	34.7	1.6	18.8	2.2	3.0	2.2	5.6	3.8
2.2	3.1	1.3	1.1	14.1	4.0	21.0	6.1	1.3	20.4
7.5	3.9	10.1	8.1	19.5	5.2	12.0	15.8	10.4	5.2
6.4	10.8	83.1	3.6	6.2	6.3	16.3	12.7	1.3	0.8
8.8	5.1	3.7	26.3	6.0	48.0	8.2	11.7	7.2	3.9
15.3	16.6	8.8	12.0	4.7	14.7	6.4	17.0	2.5	16.2

Sin organización, es difícil tener una idea de las características más importantes de los datos, qué podría ser un valor típico (o representativo), si los valores están muy concen-

```
Stem-and-leaf of FundRsng N = 60
Leaf Unit = 1.0
    0 | 0111112222333333344
    0 | 55556666666778888
    1 | 0001222244
    1 | 55666789
    2 | 01
    2 | 6
    3 | 4
    3 |
    4 |
    4 | 8
    5 |
    5 |
    6 |
    6 |
    7 |
    7 |
    8 | 3
```

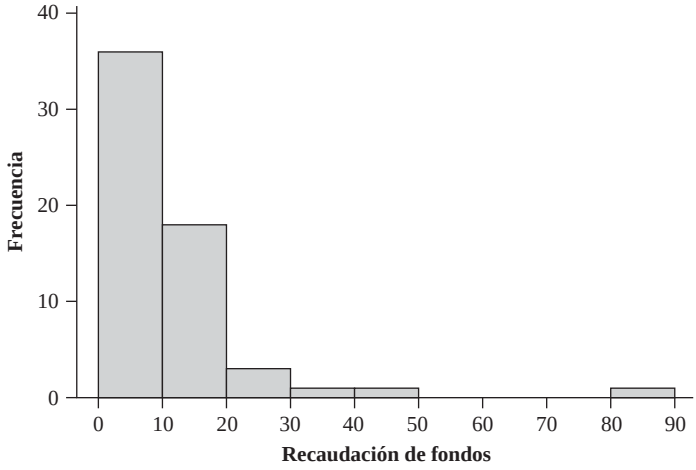


Figura 1.1 Gráfica de tallos y hojas (truncada a diez dígitos) de Minitab e histograma para los datos del porcentaje de recaudación de fondos de caridad

trados en torno a un valor típico o bastante dispersos, si existan brechas en los datos, qué porcentajes de los valores son menores a 20%, y así sucesivamente. La figura 1.1 muestra lo que se conoce como *gráfica de tallo y hojas* de los datos, así como también un *histograma*. En la sección 1.2 se discutirá la construcción e interpretación de estos resúmenes gráficos; por el momento se espera que se vea cómo los porcentajes están distribuidos sobre el rango de valores de 0 a 100. Es claro que la mayoría de las organizaciones de caridad en el ejemplo gastan menos de 20% en recaudar fondos y sólo unos pequeños porcentajes podrían ser vistos más allá del límite de una práctica sensible. ■

Después de haber obtenido una muestra de una población, un investigador con frecuencia desearía utilizar la información muestral para sacar algún tipo de conclusión (hacer una inferencia de alguna clase) con respecto a la población. Es decir, la muestra es un medio para llegar a un fin en lugar de un fin por sí misma. Las técnicas para generalizar a partir de una muestra hasta una población se congregan dentro de la rama de la disciplina llamada **estadística inferencial.**

Ejemplo 1.2 Las investigaciones de resistencia de materiales constituyen una rica área de aplicación de métodos estadísticos. El artículo "Effects of Aggregates and Microfillers on the Flexural Properties of Concrete" (*Magazine of Concrete Research*, 1997: 81–98) reportó sobre un estudio de propiedades de resistencia de concreto de alto desempeño obtenido con el uso de superplastificantes y ciertos aglomerantes. La resistencia a la compresión de dicho concreto previamente había sido investigada, pero no se sabía mucho sobre la resistencia a la flexión (una medida de la capacidad de resistir fallas por flexión). Los datos anexos sobre resistencia a la flexión (en megapascales, MPa, donde 1 Pa (pascal) = 1.45×10^{-4} lb/pulg^2) aparecieron en el artículo citado:

5.9	7.2	7.3	6.3	8.1	6.8	7.0	7.6	6.8	6.5	7.0	6.3	7.9	9.0
8.2	8.7	7.8	9.7	7.4	7.7	9.7	7.8	7.7	11.6	11.3	11.8	10.7	

Supóngase que se desea *estimar* el valor promedio de resistencia a la flexión de todas las vigas que pudieran ser fabricadas de esta manera (si se conceptualiza una población de todas esas vigas, se trata de estimar la media poblacional). Se puede demostrar que, con un alto grado de confianza, la resistencia media de la población se encuentra entre 7.48 MPa y 8.80 MPa; esto se llama *intervalo de confianza* o *estimación de intervalo*. Alternativamente, se podrían utilizar estos datos para predecir la resistencia a la flexión de una *sola* viga de este tipo. Con un alto grado de confianza, la resistencia de una sola viga excederá de 7.35 MPa; el número 7.35 se conoce como *límite de predicción inferior*. ■

El objetivo principal de este libro es presentar e ilustrar métodos de estadística inferencial que son útiles en el trabajo científico. Los tipos más importantes de procedimientos inferenciales, estimación puntual, comprobación de hipótesis y estimación por medio de intervalos de confianza, se introducen en los capítulos 6 a 8 y luego se utilizan escenarios más complicados en los capítulos 9 a 16. El resto de este capítulo presenta métodos de estadística descriptiva que se utilizan mucho en el desarrollo de inferencia.

Los capítulos 2 a 5 presentan material de la disciplina de probabilidad. Este material finalmente tiende un puente entre las técnicas descriptivas e inferenciales. El dominio de la probabilidad permite entender mejor cómo se desarrollan y utilizan los procedimientos inferenciales, cómo las conclusiones estadísticas pueden ser traducidas al lenguaje diario e interpretadas y cuándo y dónde pueden ocurrir errores al aplicar los métodos. La probabilidad y estadística se ocupan de cuestiones que implican poblaciones y muestras, pero lo hacen de una "manera inversa" una con respecto a la otra.

En un problema de probabilidad, se supone que las propiedades de la población estudiada son conocidas (p. ej., en una población numérica, se puede suponer una cierta distribución especificada de valores de la población) y se pueden plantear y responder preguntas con respecto a una muestra tomada de una población. En un problema de esta-

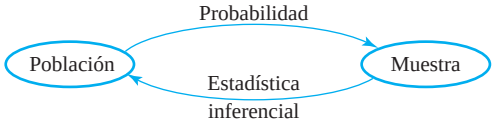


Figura 1.2 Relación entre probabilidad y estadística inferencial

dística, el experimentador dispone de las características de una muestra y esta información le permite sacar conclusiones con respecto a la población. La relación entre las dos disciplinas se resume diciendo que la probabilidad discurre de la población a la muestra (razonamiento deductivo), mientras que la estadística inferencial discurre de la muestra a la población (razonamiento inductivo). Esto se ilustra en la figura 1.2.

Antes de que se pueda entender lo que una muestra particular pueda decir sobre la población, primero se deberá entender la incertidumbre asociada con la toma de una muestra de una población dada. Por eso se estudia la probabilidad antes que la estadística.

Ejemplo 1.3 Como un ejemplo del enfoque contrastante de la probabilidad y la estadística inferencial, considere el uso que los conductores hacen de los cinturones de seguridad manuales de regazo en carros equipados con sistemas de cinturones de hombro automáticos. (El artículo "Automobile Seat Belts: Usage Patterns in Automatic Belt Systems", *Human Factors*, 1998: 126–135, resume datos de uso.) Se podría suponer que probablemente el 50% de todos los conductores de carros equipados de esta manera en cierta área metropolitana utilizan de manera regular su cinturón de regazo (una suposición sobre la población), así que se podría preguntar, "¿qué tan probable es que una muestra de 100 conductores incluirá por lo menos 70 que regularmente utilicen su cinturón de regazo?" o "¿cuántos de los conductores en una muestra de tamaño 100 se puede esperar que utilicen con regularidad su cinturón de regazo?". Por otra parte, en estadística inferencial se dispone de información sobre la muestra; por ejemplo, una muestra de 100 conductores de tales vehículos reveló que 65 utilizan con regularidad su cinturón de regazo. Se podría entonces preguntar: "¿Proporciona esto evidencia sustancial para concluir que más del 50% de todos los conductores en esta área utilizan con regularidad su cinturón de regazo?" En el último escenario se intenta utilizar la información sobre la muestra para responder una pregunta sobre la estructura de toda la población de la cual se seleccionó la muestra. ■

En el ejemplo del cinturón de regazo, la población está bien definida y concreta: todos los conductores de carros equipados de una cierta manera en un área metropolitana particular. En el ejemplo 1.2, sin embargo, las mediciones de resistencia vienen de una muestra de vigas prototipo que no tuvieron que seleccionarse de una población existente. En su lugar, conviene pensar en la población como compuesta de todas las posibles mediciones de resistencia que se podrían hacer en condiciones experimentales similares. Tal población se conoce como **población conceptual** o **hipotética.** Existen varias situaciones en las cuales las preguntas encajan en el marco de referencia de la estadística inferencial al conceptualizar una población.

El ámbito de la estadística moderna

Actualmente, la metodología estadística es empleada por investigadores en prácticamente todas las disciplinas, incluyendo algunas áreas como

- biología molecular (en el análisis de datos de microarreglos)
- ecología (en la descripción cuantitativa de cómo individuos de varias poblaciones de plantas y animales están distribuidos espacialmente)

- ingeniería de materiales (en el estudio de propiedades de varios tratamientos para retardar la corrosión)
- marketing (en el desarrollo de estudios de mercado y estrategias para la comercialización de nuevos productos)
- salud pública (en la identificación de fuentes de enfermedades y sus formas de tratamiento)
- ingeniería civil (en la evaluación de los efectos de los esfuerzos en los elementos estructurales y los impactos del flujo del tránsito de vehículos en las comunidades)

A medida que se progrese en el libro, se encontrará un amplio espectro de escenarios diferentes en los ejemplos y ejercicios que ilustran la aplicación de técnicas de probabilidad y estadística. Muchos de estos escenarios involucran datos u otros materiales extraídos de artículos de ingeniería y revistas de ciencia. Los métodos aquí presentados convierten herramientas establecidas y confiables en el arsenal de todo aquel que trabaja con datos. Mientras tanto, los estadísticos continúan desarrollando nuevos modelos para describir aleatoriedad, incertidumbre y una metodología nueva para el análisis de datos. Como evidencia de los continuos esfuerzos creativos en la comunidad estadística, existen títulos y cápsulas con descripciones de artículos publicados recientemente en revistas de estadística (*Journal of the American Statistical Association* y los *Annals of Applied Statistics*, cuyas siglas son *JASA* y *AAS*, respectivamente, dos de las revistas más importantes en esta disciplina):

- "Modeling Spatiotemporal Forest Health Monitoring Data"(*JASA*, 2009: 899–911): Sistemas de vigilancia de la salud forestal se crearon en toda Europa en la década de 1980 en respuesta a la preocupación por la desaparición de los bosques relacionada con la contaminación del aire y han continuado operando con un enfoque más reciente sobre las amenazas del cambio climático y el aumento de los niveles de ozono. Los autores desarrollan una descripción cuantitativa de la defoliación de copas, un indicador de la salud de los árboles.
- "Active Learning Through Sequential Design, with Applications to the Detection of Money Laundering" (*JASA*, 2009: 969–981): El lavado de dinero consiste en ocultar el origen de los fondos obtenidos a través de actividades ilegales. El enorme número de transacciones que ocurren a diario en las instituciones financieras dificulta la detección del lavado de capitales. El planteamiento más común ha sido extraer un resumen de diversas cantidades de la historia de las transacciones y llevar a cabo una investigación de mucho tiempo de actividades sospechosas. El artículo propone un método estadístico más eficiente e ilustra su uso en un caso de estudio.
- "Robust Internal Benchmarking and False Discovery Rates for Detecting Racial Bias in Police Stops" (*JASA*, 2009:661–668): Alegatos de las acciones policiales atribuidas al menos en parte a los prejuicios raciales se han convertido en un tema polémico en muchas comunidades. En este artículo se propone un nuevo método que está diseñado para reducir el riesgo de marcar un número sustancial de "falsos positivos" (personas falsamente identificadas como la manifestación de un sesgo). El método se aplicó a los datos de 500,000 peatones detenidos en la ciudad de Nueva York en 2006; de los 3000 agentes que participan regularmente en la detención de peatones, 15 fueron identificados por haber detenido una fracción mucho mayor de negros e hispanos de lo que podría predecirse en ausencia del sesgo.
- "Records in Athletics Through Extreme Value Theory"(*JASA*, 2008:1382–1391): El documento se centra en el modelado de los extremos relacionados con récords mundiales en atletismo. Los autores comienzan planteando dos cuestiones: (1) ¿Cuál es el último récord mundial en un evento específico (por ejemplo, el salto de altura para las mujeres)? y (2) ¿Cuán "bueno" es el actual récord mundial y cómo es la calidad de los actuales récords del mundo al comparar los diferentes eventos? Se considera

un total de 28 eventos (8 carreras, 3 lanzamientos, y 3 saltos para los hombres y mujeres). Por ejemplo, una conclusión es que el récord masculino de maratón sólo se ha reducido 20 segundos, pero el que registran las mujeres actualmente en el maratón es casi 5 minutos más de lo que en última instancia se puede lograr. La metodología también tiene aplicaciones a problemas tales como asegurar que las pistas de aterrizaje sean lo suficientemente largas o que los diques en Holanda sean lo suficientemente altos.

- "Analysis of Episodic Data with Application to Recurrent Pulmonary Exacerbations in Cystic Fibrosis Patients" (*JASA*, 2008: 498–510): El análisis de los eventos médicos recurrentes, tales como dolores de cabeza por migraña, deben tener en cuenta no sólo cuándo tales eventos aparecen por primera vez, sino también su duración, la gran duración de los episodios pueden contener información importante acerca de la gravedad de la enfermedad, los costos médicos asociados y la calidad de vida. El artículo propone una técnica que resume tanto la frecuencia de los episodios y la duración de los mismos y permite que la ocurrencia del episodio y las características de los efectos puedan variar con el tiempo. La técnica se aplica a los datos sobre pacientes con fibrosis quística (FQ es un trastorno genético grave que afecta las glándulas sudoríparas y otras).
- "Prediction of Remaining Life of Power Transformers Based on Left Truncated and Right Censored Lifetime Data" (*AAS*, 2009: 857–879): Hay aproximadamente 150,000 transformadores de transmisión de energía de alta tensión en Estados Unidos. Fallas inesperadas pueden causar grandes pérdidas económicas, por lo que es importante contar con las predicciones de vida restante. Los datos pertinentes pueden ser complicados, porque la vida útil de algunos transformadores se extiende por varias décadas durante las cuales los registros no eran necesariamente completos. En particular, los autores del artículo utilizan datos de una empresa de energía que comenzó a llevar registros detallados en 1980. Sin embargo, algunos transformadores se habían instalado antes de enero 1 de 1980 y todavía estaban en servicio después de esa fecha ("truncamiento a la izquierda" de datos), mientras que otras unidades estaban aún en servicio en el momento de la investigación, por lo que su vida completa no está disponible ("truncamiento a la derecha" de datos). El artículo describe los diversos procedimientos para la obtención de un intervalo de valores posibles (un *intervalo de predicción*) para toda la vida restante y el número acumulado de fallas en un periodo especificado.
- "The BARISTA: A Model for Bid Arrivals in Online Auctions" (*AAS*, 2007: 412–441): Las subastas en línea como las de eBay y uBid a menudo tienen características que las diferencian de las subastas tradicionales. Una diferencia muy importante es que el número de oferentes en el comienzo de muchas subastas tradicionales es fijo, mientras que en las subastas en línea este número y el número de ofertas resultantes no está predeterminado. El artículo propone una nueva BARISTA (Bid ARrivals In STAges) modelo para describir la forma en que llegan las ofertas en línea. El modelo permite hacer de manera intensa una oferta más alta al comienzo de la subasta y también cuando ésta llega a su fin. Varias propiedades del modelo son investigadas y validadas con datos de las subastas de eBay.com para asistentes personales Palm M515, juegos de Microsoft Xbox y relojes Cartier.
- "Statistical Challenges in the Analysis of Cosmic Microwave Background Radiation" (*AAS*, 2009: 61–95): El fondo cósmico de microondas (CMB, por sus siglas en inglés) es una fuente importante de información sobre la historia temprana del universo. Su nivel de radiación es uniforme, e instrumentos extremadamente delicados han sido desarrollados para medir las fluctuaciones. Los autores proporcionan una revisión de las cuestiones estadísticas del CMB con el análisis de datos, también dan muchos ejemplos de la aplicación de procedimientos estadísticos a los datos obtenidos de una reciente misión del satélite de la NASA, la *Wilkinson Microwave Anisotropy Probe*.

La información estadística aparece ahora con mayor frecuencia en los medios populares y en ocasiones el centro de atención se enfoca en los estadísticos. Por ejemplo, el 23 de noviembre de 2009, el *New York Times* reportó en un artículo "Behind Cancer Guidelines, Quest for Data" que la nueva ciencia para la investigación del cáncer y métodos más sofisticados para el análisis de los datos hecho por los servicios preventivos de EE.UU. impulsó un grupo de trabajo para reexaminar las directrices de cómo las mujeres, frecuentemente de mediana edad en adelante, deben hacerse mamografías. El panel formó seis grupos independientes para hacer modelos estadísticos. El resultado fue un nuevo conjunto de conclusiones, incluida la afirmación de que las mamografías cada dos años son tan beneficiosas como las mamografías anuales para las pacientes, pero con sólo la mitad del riesgo de sufrir daños. Donald Berry, un bioestadístico muy prominente, fue citado diciendo que estaba gratamente sorprendido de que el grupo de trabajo tomara la nueva investigación en serio para la formulación de sus recomendaciones. El informe del grupo de trabajo ha generado mucha controversia entre las organizaciones del cáncer, los políticos y las propias mujeres.

Es nuestra esperanza que usted esté cada vez más convencido de la importancia y la pertinencia de la disciplina de la estadística, así como de excavar más profundamente en el libro y el tema. Esperamos que se le motive lo suficiente como para querer continuar con su aprendizaje de la estadística más allá de su curso actual.

Estudios enumerativos contra analíticos

W. E. Deming, estadístico estadounidense muy influyente quien fue una fuerza propulsora en la revolución de calidad de Japón durante las décadas de 1950 y 1960, introdujo la distinción entre *estudios enumerativos* y *estudios analíticos*. En los primeros, el interés se enfoca en un conjunto de individuos u objetos finito, identificable y no cambiante que conforman una población. Un *marco de muestreo*, es decir, una lista de los individuos u objetos que tienen que ser muestreados, está disponible para un investigador o puede ser construido. Por ejemplo, el marco se podría componer de todas las firmas incluidas en una petición para calificar cierta iniciativa para las boletas de votación en una elección próxima; por lo general se elige una muestra para indagar si el número de firmas *válidas* sobrepasa un valor especificado. Como otro ejemplo, el marco puede contener números de serie de todos los hornos fabricados por una compañía particular durante cierto lapso de tiempo; se puede seleccionar una muestra para inferir algo sobre la duración promedio de estas unidades. El uso de métodos inferenciales presentados en este libro es razonablemente no controversial en tales escenarios (aun cuando los estadísticos continúan argumentando sobre qué métodos particulares deben ser utilizados).

Un estudio analítico se define ampliamente como uno que no es de naturaleza enumerativa. Tales estudios a menudo se realizan con el objetivo de mejorar un producto futuro al actuar sobre un proceso de cierta clase (p. ej., recalibrar equipo o ajustar el nivel de alguna sustancia tal como la cantidad de un catalizador). A menudo se obtienen datos sólo sobre un proceso existente, uno que puede diferir en aspectos importantes del proceso futuro. No existe por lo tanto un marco de muestreo que enliste los individuos u objetos de interés. Por ejemplo, una muestra de cinco turbinas con un nuevo diseño puede ser fabricada y probada para investigar su eficiencia. Estas cinco podrían ser consideradas como una muestra de la población conceptual de todos los prototipos que podrían ser fabricados experimentalmente en condiciones similares, pero *no* necesariamente representativas de la población de las unidades fabricadas una vez que la producción futura esté en proceso. Los métodos para utilizar la información sobre muestras para sacar conclusiones sobre unidades de producción futuras pueden ser problemáticos. Se deberá llamar a alguien con los conocimientos necesarios en el área del diseño e ingeniería de turbinas (o de cualquier otra área pertinente) para que juzgue si tal extrapolación es sensible. Una buena exposición de estos temas se encuentra en el artículo "Assumptions for Statistical Inference", de Gerald Hahn y William Meeker (*The American Statistician*, 1993: 1–11).

Recopilación de datos

La estadística se ocupa no sólo de la organización y análisis de datos una vez que han sido recopilados sino también del desarrollo de técnicas de recopilación de datos. Si éstos no son apropiadamente recopilados, un investigador no puede ser capaz de responder las preguntas consideradas con un razonable grado de confianza. Un problema común es que la población objetivo, aquella sobre la cual se van a sacar conclusiones, puede ser diferente de la población realmente muestreada. Por ejemplo, a los publicistas les gustaría contar con varias clases de información sobre los hábitos de ver televisión de sus clientes potenciales. La información más sistemática de esta clase proviene de colocar dispositivos de monitoreo en un pequeño número de casas a través de Estados Unidos. Se ha conjeturado que la colocación de semejantes dispositivos por sí misma modifica el comportamiento del televidente, de modo que las características de la muestra pueden ser diferentes de aquellas de la población objetivo.

Cuando la recopilación de datos implica seleccionar individuos u objetos de un marco, el método más simple para garantizar una selección representativa es tomar una *muestra aleatoria simple*. Ésta es una para la cual cualquier subconjunto particular del tamaño especificado (p. ej., una muestra de tamaño 100) tiene la misma oportunidad de ser seleccionada. Por ejemplo, si el marco se compone de 1,000,000 de números de serie, los números 1, 2, . . . , hasta 1,000,000 podrían ser anotados en trozos idénticos de papel. Después de colocarlos en una caja y mezclarlos perfectamente, se sacan uno por uno hasta que se obtenga el tamaño de muestra requisito. De manera alternativa (y mucho más preferible), se podría utilizar una tabla de números aleatorios o un generador de números aleatorios de computadora.

En ocasiones se pueden utilizar métodos de muestreo alternativos para facilitar el proceso de selección, a fin de obtener información extra o para incrementar el grado de confianza en conclusiones. Un método como ése, el *muestreo estratificado*, implica separar las unidades de la población en grupos no traslapantes y tomar una muestra de cada uno. Por ejemplo, un fabricante de reproductores de DVD podría desear información sobre la satisfacción del cliente para unidades producidas durante el año previo. Si tres modelos diferentes fueran fabricados y vendidos, se podría seleccionar una muestra distinta de cada uno de los estratos correspondientes. Esto daría información sobre los tres modelos y garantizaría que ningún modelo estuviera sobre o subrepresentado en toda la muestra.

Con frecuencia, se obtiene una muestra de "conveniencia" seleccionando individuos u objetos sin aleatorización sistemática. Por ejemplo, un conjunto de ladrillos puede ser apilado de tal modo que sea extremadamente difícil seleccionar a los que se encuentran en el centro. Si los ladrillos localizados en la parte superior y a los lados de la pila fueran de algún modo diferentes a los demás, los datos muestrales resultantes no representarían la población. A menudo un investigador supondrá que tal muestra de conveniencia representa en forma aproximada una muestra aleatoria, en cuyo caso el repertorio de métodos inferenciales de un estadístico puede ser utilizado; sin embargo, ésta es una cuestión de criterio. La mayoría de los métodos aquí analizados se basan en una variación del muestreo aleatorio simple descrito en el capítulo 5.

Los ingenieros y científicos a menudo reúnen datos realizando alguna clase de experimento. Esto puede implicar cómo asignar varios tratamientos diferentes (tales como fertilizantes o recubrimientos anticorrosivos) a las varias unidades experimentales (parcelas o tramos de tubería). Por otra parte, un investigador puede variar sistemáticamente los niveles o categorías de ciertos factores (p. ej., presión o tipo de material aislante) y observar el efecto en alguna variable de respuesta (tal como rendimiento de un proceso de producción).

Ejemplo 1.4 Un artículo en el *New York Times* (27 de enero de 1987) reportó que el riesgo de sufrir un ataque cardiaco podría ser reducido tomando aspirina. Esta conclusión se basó en un experimento diseñado que incluía tanto un grupo de control de individuos que tomaron un placebo que tenía la apariencia de aspirina pero que se sabía era inerte y un grupo de

tratamiento que tomó aspirina de acuerdo con un régimen específico. Los sujetos fueron asignados al azar a los grupos para protegerlos contra cualquier prejuicio de modo que se pudieran utilizar métodos basados en la probabilidad para analizar los datos. De los 11,034 individuos en el grupo de control, 189 experimentaron subsecuentemente ataques cardiacos, mientras que sólo 104 de los 11,037 en el grupo de aspirina sufrieron un ataque cardiaco. La tasa de incidencia de ataques cardiacos en el grupo de tratamiento fue de sólo aproximadamente la mitad de aquella en el grupo de control. Una posible explicación de este resultado es la variación de la probabilidad, que la aspirina en realidad no tiene el efecto deseado y la diferencia observada es sólo una variación típica del mismo modo que el lanzamiento al aire de dos monedas idénticas por lo general produciría diferentes cantidades de águilas. No obstante, en este caso, los métodos inferenciales sugieren que la variación de la probabilidad por sí misma no puede explicar en forma adecuada la magnitud de la diferencia observada. ■

Ejemplo 1.5 Un ingeniero desea investigar los efectos tanto del tipo de adhesivo como del material conductor en la fuerza adhesiva cuando se monta un circuito integrado (CI) sobre cierto sustrato. Se consideraron dos tipos de adhesivo y dos materiales conductores. Se realizaron dos observaciones por cada combinación de tipo de adhesivo/material conductor y se obtuvieron los datos anexos.

Tipo de adhesivo	Material conductor	Fuerza adhesiva observada	Promedio
1	1	82, 77	79.5
1	2	75, 87	81.0
2	1	84, 80	82.0
2	2	78, 90	84.0

Las fuerzas adhesivas promedio resultantes se ilustran en la figura 1.3. Parece que el adhesivo tipo 2 mejora la fuerza adhesiva en comparación con el tipo 1 en aproximadamente la misma cantidad siempre que se utiliza uno de los materiales conductores, con la combinación 2, 2 como la mejor. De nuevo se pueden utilizar métodos inferenciales para juzgar si estos efectos son reales o simplemente se deben a la variación de la probabilidad.

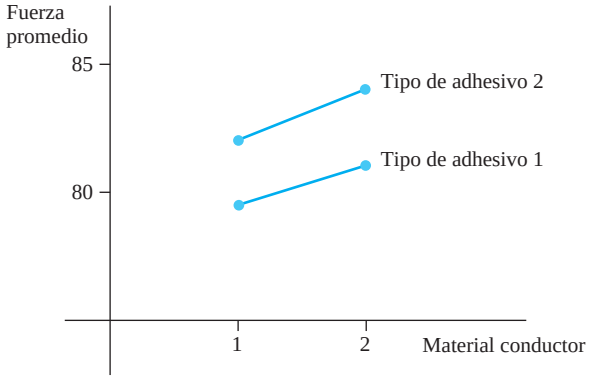


Figura 1.3 Fuerzas adhesivas promedio en el ejemplo 1.5

Supóngase además que se consideran dos tiempos de curado y también dos tipos de posrecubrimientos de los circuitos integrados. Existen entonces $2 \cdot 2 \cdot 2 \cdot 2 = 16$ combinaciones de estos cuatro factores y es posible que el ingeniero no disponga de suficientes recursos para hacer incluso una observación sencilla para cada una de estas combinaciones. En el capítulo 11 se verá cómo la selección cuidadosa de una fracción de estas posibilidades usualmente dará la información deseada. ■

EJERCICIOS Sección 1.1 (1–9)

1. Dé una posible muestra de tamaño 4 de cada una de las siguientes poblaciones.
- **a.** Todos los periódicos publicados en Estados Unidos
- **b.** Todas las compañías listadas en la Bolsa de Valores de Nueva York.
- **c.** Todos los estudiantes en su colegio o universidad.
- **d.** Todas las calificaciones promedio de los estudiantes en su colegio o universidad.

2. Para cada una de las siguientes poblaciones hipotéticas, dé una muestra posible de tamaño 4:
- **a.** Todas las distancias que podrían resultar cuando usted lanza un balón de futbol americano.
- **b.** Las longitudes de las páginas de libros publicados de aquí a 5 años.
- **c.** Todas las mediciones de intensidades posibles de terremotos (escala de Richter) que pudieran registrarse en California durante el siguiente año.
- **d.** Todos los posibles rendimientos (en gramos) de una cierta reacción química realizada en un laboratorio.

3. Considere la población compuesta de todas las computadoras de una cierta marca y modelo y enfóquese en si una computadora necesita servicio mientras se encuentra dentro de la garantía.
- **a.** Plantee varias preguntas de probabilidad con base en la selección de 100 de esas computadoras.
- **b.** ¿Qué pregunta de estadística inferencial podría ser respondida determinando el número de dichas computadoras en una muestra de tamaño 100 que requieren servicio de garantía?

4. **a.** Dé tres ejemplos diferentes de poblaciones concretas y tres ejemplos distintos de poblaciones hipotéticas.
- **b.** Por cada una de sus poblaciones concretas e hipotéticas, dé un ejemplo de una pregunta de probabilidad y un ejemplo de pregunta de estadística inferencial.

5. Muchas universidades y colegios han instituido programas de instrucción suplementaria (IS), en los cuales un facilitador regularmente se reúne con un pequeño grupo de estudiantes inscritos en el curso para promover discusiones sobre el material incluido en el curso y mejorar el dominio de la materia. Suponga que los estudiantes inscritos en un largo curso de estadística (¿de qué más?) se dividen al azar en un grupo de control que no participará en la instrucción suplementaria y en un grupo de tratamiento que sí participará. Al final del curso, se determina la calificación total de cada estudiante en el curso.
- **a.** ¿Son las calificaciones del grupo IS una muestra de una población existente? De ser así, ¿cuál es? De no ser así, ¿cuál es la población conceptual pertinente?
- **b.** ¿Cuál piensa que es la ventaja de dividir al azar a los estudiantes en los dos grupos en lugar de permitir que cada estudiante elija el grupo al que desea unirse?
- **c.** ¿Por qué los investigadores no pusieron a todos los estudiantes en el grupo de tratamiento? *Nota*: el artículo “Supplemental Instruction: An Effective Component of Student Affairs Programming” (*J. of College Student Devel*., 1997:577–586) discute el análisis de datos de varios programas de instrucción suplementaria.

6. El sistema de la Universidad Estatal de California (CSU, por sus siglas en inglés) consta de 23 campus universitarios, desde la Estatal de San Diego en el sur hasta la Estatal Humboldt cerca de la frontera con Oregon. Un administrador de CSU desea hacer una inferencia sobre la distancia promedio entre la ciudad natal de los estudiantes y sus campus universitarios. Describa y discuta varios diferentes métodos de muestreo que pudieran ser empleados. ¿Éste sería un estudio enumerativo o un estudio analítico? Explique su razonamiento.

7. Cierta ciudad se divide naturalmente en diez distritos. ¿Cómo podría seleccionar un valuador de bienes raíces una muestra de casas unifamiliares que pudiera ser utilizada como base para desarrollar una ecuación para predecir el valor estimado a partir de características tales como antigüedad, tamaño, número de baños, distancia a la escuela más cercana y así sucesivamente? ¿El estudio es enumerativo o analítico?

8. La cantidad de flujo a través de una válvula solenoide en el sistema de control de emisiones de un automóvil es una característica importante. Se realizó un experimento para estudiar cómo la velocidad de flujo dependía de tres factores: la longitud de la armadura, la fuerza del resorte y la profundidad de la bobina. Se eligieron dos niveles diferentes (alto y bajo) de cada factor y se realizó una sola observación del flujo por cada combinación de niveles.
- **a.** ¿De cuántas observaciones consistió el conjunto de datos resultante?
- **b.** ¿Este estudio es enumerativo o analítico? Explique su razonamiento.

9. En un famoso experimento realizado en 1882, Michelson y Newcomb obtuvieron 66 observaciones del tiempo que requería la luz para viajar entre dos lugares en Washington, D.C. Algunas de las mediciones (codificadas en cierta manera) fueron, 31, 23, 32, 36, –2, 26, 27 y 31.
- **a.** ¿Por qué no son idénticas estas mediciones?
- **b.** ¿Es éste un estudio enumerativo? ¿Por qué sí o por qué no?

1.2 Métodos pictóricos y tabulares en la estadística descriptiva

La estadística descriptiva se divide en dos temas generales. En esta sección se considera la representación de un conjunto de datos por medio de técnicas visuales. En las secciones 1.3 y 1.4 se desarrollarán algunas medidas numéricas para conjuntos de datos. Es posible que usted ya conozca muchas técnicas visuales; tablas de frecuencia, hojas de contabili-

dad, histogramas, gráficas de pastel, gráficas de barras, diagramas de puntos y similares. Aquí se seleccionan algunas de estas técnicas que son más útiles y pertinentes para la probabilidad y estadística inferencial.

Notación

Alguna notación general facilitará la aplicación de métodos y fórmulas a una amplia variedad de problemas prácticos. El número de observaciones en una muestra única, es decir, el *tamaño de muestra*, a menudo será denotado por n, de modo que $n = 4$ para la muestra de universidades {Stanford, Iowa State, Wyoming, Rochester} y también para la muestra de lecturas de pH {6.3, 6.2, 5.9, 6.5}. Si se consideran dos muestras al mismo tiempo, m y n o n_1 y n_2 se pueden utilizar para denotar los números de observaciones. Por lo tanto si {29.7, 31.6, 30.9} y {28.7, 29.5, 29.4, 30.3} son lecturas de eficiencia térmica de dos tipos diferentes de motores diesel, entonces $m = 3$ y $n = 4$.

Dado un conjunto de datos compuesto de n observaciones de alguna variable x, entonces $x_1, x_2, x_3, \ldots, x_n$ denotarán las observaciones individuales. El subíndice no guarda ninguna relación con la magnitud de una observación particular. Por lo tanto x_1 en general no será la observación más pequeña del conjunto, ni x_n será la más grande. En muchas aplicaciones, x_1 será la primera observación realizada por el experimentador, x_2 la segunda, y así sucesivamente. La observación i-ésima del conjunto de datos será denotada por x_i.

Gráficas de tallos y hojas

Considérese un conjunto de datos numéricos $x_1, x_2, \ldots, x_n$ para el cual cada x_i se compone de por lo menos dos dígitos. Una forma rápida de obtener la representación visual informativa del conjunto de datos es construir una *gráfica de tallos y hojas*.

Pasos para construir una gráfica de tallos y hojas

1. Seleccione uno o más de los primeros dígitos para los valores de tallo. Los segundos dígitos se convierten en hojas.
2. Enumere los posibles valores de tallos en una columna vertical.
3. Anote la hoja para cada observación junto al correspondiente valor de tallo.
4. Indique las unidades para tallos y hojas en algún lugar de la gráfica.

Si el conjunto de datos se compone de calificaciones de exámenes, cada uno entre 0 y 100, la calificación de 83 tendría un tallo de 8 y una hoja de 3. Para un conjunto de datos de eficiencias de consumo de combustible de automóviles (mpg), todos entre 8.1 y 47.8, se podría utilizar el dígito de las decenas como el tallo, así que 32.6 tendría entonces una hoja de 2.6. En general, se recomienda una gráfica basada en tallos entre 5 y 20.

Ejemplo 1.6 El consumo de alcohol por parte de estudiantes universitarios preocupa no sólo a la comunidad académica sino también, a causa de consecuencias potenciales de salud y seguridad, a la sociedad en su conjunto. El artículo "Health and Behavioral Consequences of Binge Drinking in College" (*J. of the Amer. Med. Assoc.*, 1994: 1672–1677 presentó un amplio estudio sobre el consumo excesivo de alcohol en universidades a través de Estados Unidos. Un episodio de parranda se definió como cinco o más tragos en fila para varones y cuatro o más para mujeres. La figura 1.4 muestra una gráfica de tallo y hojas de 140 valores de x = porcentaje de edades de los estudiantes de licenciatura bebedores. (Estos valores no aparecieron en el artículo citado, pero la ilustración concuerda con una gráfica de los datos que sí se incluyó.)

La primera hoja de la fila 2 del tallo es 1, la cual dice que 21% de los estudiantes de una de las universidades de la muestra eran bebedores. Sin la identificación de los dígitos

```
0 | 4
1 | 1345678889
2 | 1223456666777889999
3 | 011223334455566667777788899999
4 | 111222223344445566666677788888999
5 | 00111222233455666667777888899
6 | 01111244455666778
```

Tallo: dígito de las decenas
Hojas: dígito de las unidades

Figura 1.4 Gráfica de tallo y hojas para el porcentaje de bebedores en cada una de las 140 universidades

en los tallos y los dígitos en las hojas, no se sabría si la observación correspondiente al tallo 2, hoja 1 debería leerse como 21%, 2.1% o .21%.

Cuando se crea una imagen a mano, la ordenación de las hojas de la más pequeña a la más grande en cada línea puede ser tediosa. Esta ordenación contribuye poco si no se dispone de información adicional. Supóngase que las observaciones hubieran sido puestas en lista en orden alfabético por nombre de la escuela, como

16% 33% 64% 37% 31% . . .

La colocación entonces de estos valores en la gráfica en este orden haría que la fila 1 del tallo tuviera 6 como su primera hoja y el principio de la fila 3 del tallo sería

3 | 371 . . .

La gráfica sugiere que un valor típico o representativo se encuentra en la fila 4 del tallo, tal vez en el rango medio de 40%. Las observaciones no aparecen muy concentradas en torno a este valor típico, como sería el caso si todos los valores estuvieran entre 20% y 49%. Esta gráfica se eleva a una sola cresta a medida que desciende, y luego declina; no hay brechas en la gráfica. La forma de la gráfica no es perfectamente simétrica, pero en su lugar parece alargarse un poco más en la dirección de las hojas bajas que en la dirección de las hojas altas. Por último, no existen observaciones que se alejen inusualmente del grueso de los datos (ningunos *valores apartados*), como sería el caso si uno de los valores de 26% hubiera sido de 86%. La característica más sobresaliente de estos datos es que, en la mayoría de las universidades de la muestra, por lo menos una cuarta parte de los estudiantes son bebedores. El problema de beber en exceso en las universidades es mucho más extenso de lo que muchos hubieran sospechado. ■

Una gráfica de tallos y hojas da información sobre los siguientes aspectos de los datos:

- identificación de un valor típico o representativo
- grado de dispersión en torno al valor típico
- presencia de brechas en los datos
- grado de simetría en la distribución de los valores
- número y localización de crestas
- presencia de valores afuera de la gráfica

Ejemplo 1.7 La figura 1.5 presenta gráficas de tallos y hojas de una muestra aleatoria de longitudes de campos de golf (yardas) designados por *Golf Magazine* como los más desafiantes en Estados Unidos. Entre la muestra de 40 campos, el más corto es de 6433 yardas de largo y el más largo de 7280 yardas. Las longitudes parecen estar distribuidas de una manera aproximadamente uniforme dentro del rango de valores presentes en la muestra. Obsérvese que la selección de tallo en este caso de un solo dígito (6 o 7) o de tres (643, . . . , 728) produciría una gráfica no informativa, primero a causa de pocos tallos y segundo a causa de demasiados.

Los programas de computadora de estadística en general no producen gráficas con tallos de dígitos múltiples. La gráfica Minitab que aparece en la figura 1.5(b) resulta de *truncar* cada observación al borrar los dígitos uno.

64	35	64	33	70	Tallo: dígitos de millares y centenas		
65	26	27	06	83	Hojas: dígitos de decenas y unidades		
66	05	94	14				
67	90	70	00	98	70	45	13
68	90	70	73	50			
69	00	27	36	04			
70	51	05	11	40	50	22	
71	31	69	68	05	13	65	
72	80	09					

(a)

```
Tallos y hojas de yardaje N = 40
Unidad de hoja = 10

   4          64  3367
   8          65  0228
  11          66  019
  18          67  0147799
 (4)          68  5779
  18          69  0023
  14          70  012455
   8          71  013666
   2          72  08
```

(b)

Figura 1.5 Gráficas de tallos y hojas de la longitud de los campos de golf (a) hojas de dos dígitos; (b) gráfica Minitab de hojas con truncamiento a un dígito ■

Gráficas de puntos

Una gráfica de puntos es un resumen atractivo de datos numéricos cuando el conjunto de datos es razonablemente pequeño o existen pocos valores de datos distintos. Cada observación está representada por un punto sobre la ubicación correspondiente en una escala de medición horizontal. Cuando un valor ocurre más de una vez, existe un punto por cada ocurrencia y estos puntos se apilan verticalmente. Como con la gráfica de tallos y hojas, una gráfica de puntos da información sobre la localización, dispersión, extremos y brechas.

Ejemplo 1.8 Aquí hay datos sobre los créditos estado por estado para la educación superior como porcentaje de los ingresos fiscales estatales y locales para el año fiscal 2006–2007 (tomados del *Statistical Abstract of the United States*); los valores se presentan en una lista, en el orden de las abreviaturas de cada estado (AL en primer lugar, WY al final):

10.8	6.9	8.0	8.8	7.3	3.6	4.1	6.0	4.4	8.3
8.1	8.0	5.9	5.9	7.6	8.9	8.5	8.1	4.2	5.7
4.0	6.7	5.8	9.9	5.6	5.8	9.3	6.2	2.5	4.5
12.8	3.5	10.0	9.1	5.0	8.1	5.3	3.9	4.0	8.0
7.4	7.5	8.4	8.3	2.6	5.1	6.0	7.0	6.5	10.3

La figura 1.6 muestra una gráfica de puntos para los datos. La característica más llamativa es la sustancial variación de un estado a otro. El valor más grande (para Nuevo México) y los dos valores más pequeños (Nueva Hampshire y Vermont) están algo separados de la mayor parte de los datos, aunque quizá no lo suficiente para considerarlos atípicos.

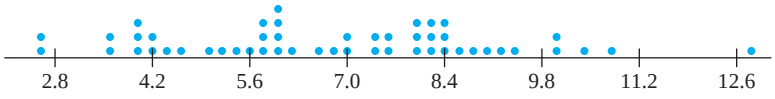


Figura 1.6 Gráfica de puntos para los datos del ejemplo 1.8

Si el número de observaciones de esfuerzo compresivo del ejemplo 1.2 hubiera consistido cuando mucho de $n = 27$ obtenidas realmente, habría sido muy tedioso construir una gráfica de puntos. La técnica siguiente es muy adecuada para situaciones como ésas.

Histogramas

Algunos datos numéricos se obtienen contando para determinar el valor de una variable (el número de citatorios de tráfico que una persona recibió durante el año pasado, el número de personas que solicitan empleo durante un lapso de tiempo particular), mientras que otros datos se obtienen tomando mediciones (peso de un individuo, tiempo de reacción a un estímulo particular). La prescripción para trazar un histograma es en general diferente en estos dos casos.

DEFINICIÓN

Una variable numérica es **discreta** si su conjunto de valores posibles es finito o además puede ser puesto en lista en una secuencia infinita (una en la cual existe un primer número, un segundo número, y así sucesivamente). Una variable numérica es **continua** si sus valores posibles abarcan un intervalo completo sobre la línea de números.

Una variable discreta x casi siempre resulta de contar, en cuyo caso los posibles valores son 0, 1, 2, 3, . . . o algún subconjunto de estos enteros. De la toma de mediciones surgen variables continuas. Por ejemplo, si x es el pH de una sustancia química, entonces en teoría x podría ser cualquier número entre 0 y 14: 7.0, 7.03, 7.032, y así sucesivamente. Desde luego, en la práctica existen limitaciones en el grado de precisión de cualquier instrumento de medición, por lo que es posible que no se pueda determinar el pH, el tiempo de reacción, la altura y la concentración con un número arbitrariamente grande de decimales. Sin embargo, desde el punto de vista de crear modelos matemáticos de distribuciones de datos, conviene imaginar un conjunto completo continuo de valores posibles.

Considérense datos compuestos de observaciones de una variable discreta x. La **frecuencia** de cualquier valor x particular es el número de veces que ocurre un valor en el conjunto de datos. La **frecuencia relativa** de un valor es la fracción o proporción de veces que ocurre el valor:

$$\text{frecuencia relativa de un valor} = \frac{\text{número de veces que ocurre el valor}}{\text{número de observaciones en el conjunto de datos}}$$

Supóngase, por ejemplo, que el conjunto de datos se compone de 200 observaciones de x = el número de cursos que un estudiante está tomando en este semestre. Si 70 de estos valores x son 3, entonces

$$\text{frecuencia del valor } x \text{ 3: } \quad 70$$

$$\text{frecuencia relativa del valor } x \text{ 3: } \quad \frac{70}{200} = .35$$

Si se multiplica una frecuencia relativa por 100 se obtiene un porcentaje; en el ejemplo de cursos universitarios, 35% de los estudiantes de la muestra están tomando tres cursos. Las frecuencias relativas, o porcentajes, por lo general interesan más que las frecuencias mismas. En teoría, las frecuencias relativas deberán sumar 1, pero en la práctica la suma puede diferir un poco de 1 por el redondeo. Una **distribución de frecuencia** es una tabla de las frecuencias o de las frecuencias relativas, o de ambas.

Construcción de un histograma para datos discretos

En primer lugar, se determinan la frecuencia y la frecuencia relativa de cada valor x. Luego se marcan los valores x posibles en una escala horizontal. Sobre cada valor, se traza un rectángulo cuya altura es la frecuencia relativa (o alternativamente, la frecuencia) de dicho valor.

Esta construcción garantiza que el *área* de cada rectángulo es proporcional a la frecuencia relativa del valor. Por lo tanto si las frecuencias relativas de $x = 1$ y $x = 5$ son .35 y .07, respectivamente, entonces el área del rectángulo sobre 1 es cinco veces el área del rectángulo sobre 5.

Ejemplo 1.9 ¿Qué tan inusual es un juego de beisbol sin hit o de un hit en las ligas mayores y cuán frecuentemente un equipo pega más de 10, 15 o incluso 20 hits? La tabla 1.1 es una distribución de frecuencia del número de hits por equipo y por juego de todos los juegos de nueve episodios que se jugaron entre 1989 y 1993.

Tabla 1.1 Distribución de frecuencia de hits en juegos de nueve entradas

Hits/juego	Número de juegos	Frecuencia relativa	Hits/juego	Número de juegos	Frecuencia relativa
0	20	.0010	14	569	.0294
1	72	.0037	15	393	.0203
2	209	.0108	16	253	.0131
3	527	.0272	17	171	.0088
4	1048	.0541	18	97	.0050
5	1457	.0752	19	53	.0027
6	1988	.1026	20	31	.0016
7	2256	.1164	21	19	.0010
8	2403	.1240	22	13	.0007
9	2256	.1164	23	5	.0003
10	1967	.1015	24	1	.0001
11	1509	.0779	25	0	.0000
12	1230	.0635	26	1	.0001
13	834	.0430	27	1	.0001
				19,383	1.0005

El histograma correspondiente en la figura 1.7 se eleva suavemente hasta una sola cresta y luego declina. El histograma se extiende un poco más hacia la derecha (hacia valores grandes) que hacia la izquierda, un poco "asimétrico positivo".

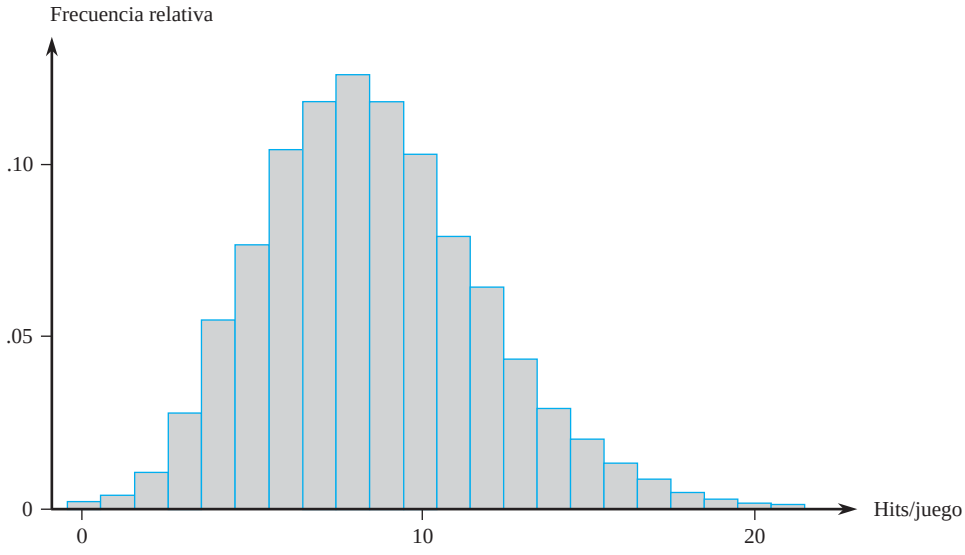


Figura 1.7 Histograma del número de hits por juego de nueve entradas

Con la información tabulada o con el histograma mismo, se puede determinar lo siguiente:

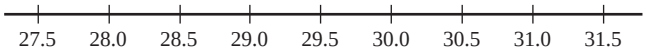
$$\begin{aligned}\text{proporción de juegos de dos hits a lo sumo} &= \begin{matrix}\text{frecuencia} \\ \text{relativa} \\ \text{para } x = 0\end{matrix} + \begin{matrix}\text{frecuencia} \\ \text{relativa} \\ \text{para } x = 1\end{matrix} + \begin{matrix}\text{frecuencia} \\ \text{relativa} \\ \text{para } x = 2\end{matrix} \\ &= .0010 + .0037 + .0108 = .0155\end{aligned}$$

Asimismo,

$$\text{proporción de juegos con entre 5 y 10 hits (inclusive)} = .0752 + .1026 + \cdots + .1015 = .6361$$

Esto es, aproximadamente 64% de todos estos juegos fueron de entre 5 y 10 hits (inclusive) ■

La construcción de un histograma para datos continuos (mediciones) implica subdividir el eje de medición en un número adecuado de **intervalos de clase** o **clases**, de tal suerte que cada observación quede contenida en exactamente una clase. Supóngase, por ejemplo, que se hacen 50 observaciones de $x =$ eficiencia de consumo de combustible de un automóvil (mpg), la más pequeña de las cuales es 27.8 y la más grande 31.4. Entonces se podrían utilizar los límites de clase 27.5, 28.0, 28.5, . . . , y 31.5 como se muestra a continuación:



Una dificultad potencial es que de vez en cuando una observación está en un límite de clase así que por consiguiente no cae en exactamente un intervalo, por ejemplo, 29.0. Una forma de habérselas con este problema es utilizar límites como 27.55, 28.05, . . . , 31.55. La adición de centésimas a los límites de clase evita que las observaciones queden en los límites resultantes. Otro método es utilizar las clases $27.5 - < 28.0$, $28.0 - < 28.5$, . . . , $31.0 - < 31.5$. En ese caso 29.0 queda en la clase $29.0 - < 29.5$ y no en la clase $28.5 - < 29.0$. En otras palabras, con esta convención, una observación que queda en el límite se coloca en el intervalo a la *derecha* del mismo. Así es como Minitab construye un histograma.

Construcción de un histograma para datos continuos: anchos de clase iguales

Se determina la frecuencia y la frecuencia relativa de cada clase. Se marcan los límites de clase sobre un eje de medición horizontal. Sobre cada intervalo de clase se traza un rectángulo cuya altura es la frecuencia relativa correspondiente (o frecuencia).

Ejemplo 1.10 Las compañías generadoras de electricidad requieren información sobre el consumo de los clientes para obtener pronósticos precisos de demandas. Investigadores de Wisconsin Power and Light determinaron el consumo de energía (en BTU) durante un periodo particular con una muestra de 90 hogares calentados con gas. Se calculó un valor de consumo ajustado como sigue:

$$\text{consumo ajustado} = \frac{\text{consumo}}{(\text{clima, en grados días})(\text{área de casa})}$$

Esto dio por resultado los datos anexos (una parte del conjunto de datos guardados FURNACE.MTW está disponible en Minitab), que se ordenaron desde el valor más pequeño al más grande.

2.97	4.00	5.20	5.56	5.94	5.98	6.35	6.62	6.72	6.78
6.80	6.85	6.94	7.15	7.16	7.23	7.29	7.62	7.62	7.69
7.73	7.87	7.93	8.00	8.26	8.29	8.37	8.47	8.54	8.58
8.61	8.67	8.69	8.81	9.07	9.27	9.37	9.43	9.52	9.58
9.60	9.76	9.82	9.83	9.83	9.84	9.96	10.04	10.21	10.28
10.28	10.30	10.35	10.36	10.40	10.49	10.50	10.64	10.95	11.09
11.12	11.21	11.29	11.43	11.62	11.70	11.70	12.16	12.19	12.28
12.31	12.62	12.69	12.71	12.91	12.92	13.11	13.38	13.42	13.43
13.47	13.60	13.96	14.24	14.35	15.12	15.24	16.06	16.90	18.26

Se permite que Minitab seleccione los intervalos de clase. La característica del histograma en la figura 1.8 que más llama la atención es su parecido a una curva en forma de campana (y por consiguiente simétrica), con el punto de simetría aproximadamente en 10.

Clase	1−<3	3−<5	5−<7	7−<9	9−<11	11−<13	13−<15	15−<17	17−<19
Frecuencia	1	1	11	21	25	17	9	4	1
Frecuencia relativa	.011	.011	.122	.233	.278	.189	.100	.044	.011

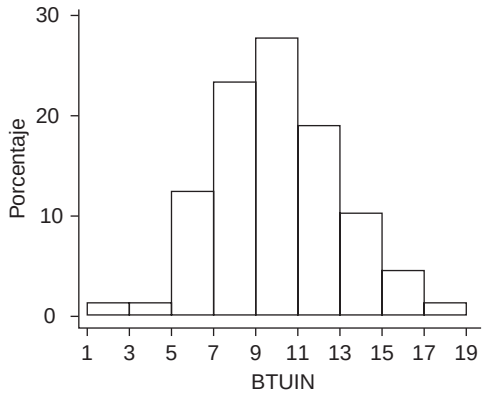


Figura 1.8 Histograma de los datos de consumo de energía del ejemplo 1.10

De acuerdo con el histograma,

$$\begin{pmatrix}\text{proporción de}\\ \text{observaciones}\\ \text{menores que 9}\end{pmatrix} \approx .01 + .01 + .12 + .23 = .37 \quad \left(\text{valor exacto} = \frac{34}{90} = .378\right)$$

La frecuencia relativa para la clase 9−<11 es aproximadamente .27, así que se estima que en forma aproximada la mitad de ésta, o .135, queda entre 9 y 10. Por lo tanto

$$\begin{pmatrix}\text{proporción de observaciones}\\ \text{menores que 10}\end{pmatrix} \approx .37 + .135 = .505 \text{ (poco más de 50\%)}$$

El valor exacto de esta proporción es 47/90 = .522. ■

No existen reglas inviolables en cuanto al número de clases o la selección de las mismas. Entre 5 y 20 será satisfactorio para la mayoría de los conjuntos de datos. En general, mientras más grande es el número de observaciones en un conjunto de datos, más clases deberán ser utilizadas. Una razonable regla empírica es

$$\text{número de clases} \approx \sqrt{\text{número de observaciones}}$$

Es posible que las clases de ancho igual no sean una opción sensible si hay regiones en la escala de medición que tienen una alta concentración de valores y otras donde los datos son muy escasos. La figura 1.9 muestra una gráfica de puntos de dicho conjunto de datos, hay alta concentración en el medio y relativamente pocas observaciones que se extienden a ambos lados. Con un pequeño número de clases de ancho igual, casi todas las observaciones quedan en exactamente una o dos de las clases. Si se utiliza un gran número de clases de ancho igual, las frecuencias de muchas clases serán cero. Una buena opción es utilizar algunos intervalos más anchos cerca de las observaciones extremas y más angostos en la región de alta concentración.

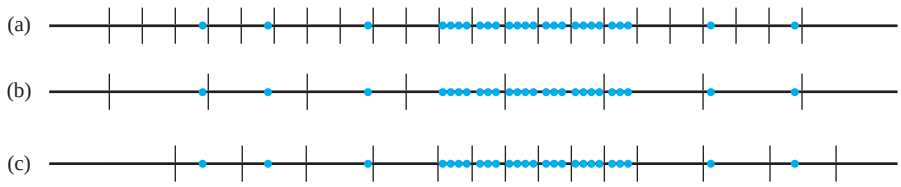


Figura 1.9 Selección de intervalos de clase para datos de "densidad variable": (a) intervalos de ancho igual muy cortos; (b) unos cuantos intervalos de ancho igual; (c) intervalos de ancho desigual

Construcción de un histograma para datos continuos: anchos de clase desiguales

Después de determinar las frecuencias y las frecuencias relativas, se calcula la altura de cada rectángulo con la fórmula

$$\text{altura del rectángulo} = \frac{\text{frecuencia relativa de la clase}}{\text{ancho de clase}}$$

Las alturas del rectángulo resultante en general se conocen como *densidades* y la escala vertical es la **escala de densidades**. Esta prescripción también funcionará cuando los anchos de clase sean iguales.

Ejemplo 1.11 La corrosión del acero de refuerzo es un problema serio en estructuras de concreto localizadas en ambientes afectados por condiciones climáticas severas. Por esa razón, los investigadores han estado estudiando el uso de barras de refuerzo hechas de un material compuesto. Se realizó un estudio para desarrollar directrices para adherir barras de refuerzo reforzadas con fibra de vidrio a concreto ("Design Recommendations for Bond of GFRP Rebars to Concrete", *J. of Structural Engr.*, 1996: 247–254). Considérense las siguientes 48 observaciones de fuerza adhesiva medida:

11.5	12.1	9.9	9.3	7.8	6.2	6.6	7.0	13.4	17.1	9.3	5.6
5.7	5.4	5.2	5.1	4.9	10.7	15.2	8.5	4.2	4.0	3.9	3.8
3.6	3.4	20.6	25.5	13.8	12.6	13.1	8.9	8.2	10.7	14.2	7.6
5.2	5.5	5.1	5.0	5.2	4.8	4.1	3.8	3.7	3.6	3.6	3.6

Clase	2−<4	4−<6	6−<8	8−<12	12−<20	20−<30
Frecuencia	9	15	5	9	8	2
Frecuencia relativa	.1875	.3125	.1042	.1875	.1667	.0417
Densidad	.094	.156	.052	.047	.021	.004

El histograma resultante aparece en la figura 1.10. La cola derecha o superior se alarga mucho más que la cola izquierda o inferior, un sustancial alejamiento de la simetría.

Cuando los anchos de clase son desiguales, si no se utiliza una escala de densidades se obtendrá una gráfica con áreas distorsionadas. Con anchos de clase iguales, el divisor es el mismo en cada cálculo de densidad y la aritmética adicional simplemente implica

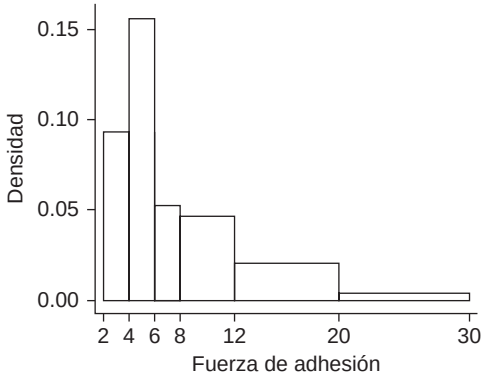


Figura 1.10 Histograma Minitab de densidad para la fuerza de adhesión del ejemplo 1.11 ■

cambiar la escala en el eje vertical (es decir, el histograma que utiliza frecuencia relativa y el que utiliza densidad tendrán exactamente la misma apariencia). Un histograma de densidad tiene una propiedad interesante. Si se multiplican ambos miembros de la fórmula para densidad por el ancho de clase se obtiene

$$\begin{aligned}\text{frecuencia relativa} &= (\text{ancho de clase})(\text{densidad})\\ &= (\text{ancho del rectángulo})(\text{altura del rectángulo})\\ &= \text{área del rectángulo}\end{aligned}$$

Es decir, *el área de cada rectángulo es la frecuencia relativa de la clase correspondiente*. Además, como la suma de frecuencias relativas debe ser 1, *el área total de todos los rectángulos en un histograma de densidad es* 1. Siempre es posible trazar un histograma de modo que el área sea igual a la frecuencia relativa (esto es cierto también para un histograma de datos discretos), simplemente se utiliza la escala de densidad. Esta propiedad desempeñará un importante papel al crear modelos de distribución en el capítulo 4.

Formas de histograma

Los histogramas se presentan en varias formas. Un histograma **unimodal** es el que se eleva a una sola cresta y luego declina. Uno **bimodal** tiene dos crestas diferentes. Puede ocurrir bimodalidad cuando el conjunto de datos se compone de observaciones de dos clases bastante diferentes de individuos u objetos. Por ejemplo, considérese un gran conjunto de datos compuesto de tiempos de manejo de automóviles que viajan entre San Luis Obispo, California y Monterey, California (sin contar el tiempo utilizado para ver puntos de interés, comer, etc.). Este histograma mostraría dos crestas, una para los carros que toman la ruta interior (aproximadamente 2.5 horas) y otra para los carros que viajan a lo largo de la costa (3.5–4 horas). Sin embargo, la bimodalidad no se presenta automáticamente en dichas situaciones. Sólo si los dos histogramas distintos están "muy alejados" en forma relativa con respecto a sus dispersiones la bimodalidad ocurrirá en el histograma de datos combinados. Por consiguiente un conjunto de datos grande compuesto de estaturas de estudiantes universitarios no producirá un histograma bimodal porque la altura típica de hombres de aproximadamente 69 pulgadas no está demasiado por encima de la altura típica de mujeres de aproximadamente 64–65 pulgadas. Se dice que un histograma con más de dos crestas es **multimodal.** Por supuesto, el número de crestas dependerá de la selección de intervalos de clase, en particular, con un pequeño número de observaciones. Mientras más grande es el número de clases, es más probable que se manifieste bimodalidad o multimodalidad.

Ejemplo 1.12 La figura 1.11(a) muestra un histograma Minitab de los pesos (en libras, lb) de los 124 jugadores que figuraban en las listas de los 49's de San Francisco y de los Patriotas de Nueva Inglaterra (equipos que al autor le gustaría ver reunidos en el Súper Tazón) el 20 de noviembre de 2009.

La figura 1.11(b) es un histograma suavizado (que en realidad se llama *densidad estimada*) de los datos del paquete de software R. El histograma y el histograma suavizado muestran tres picos diferentes; el primero a la derecha es para los linieros, el del centro corresponde al peso de los apoyadores y el pico de la izquierda es para todos los demás jugadores (receptores abiertos, mariscal de campo, etc.).

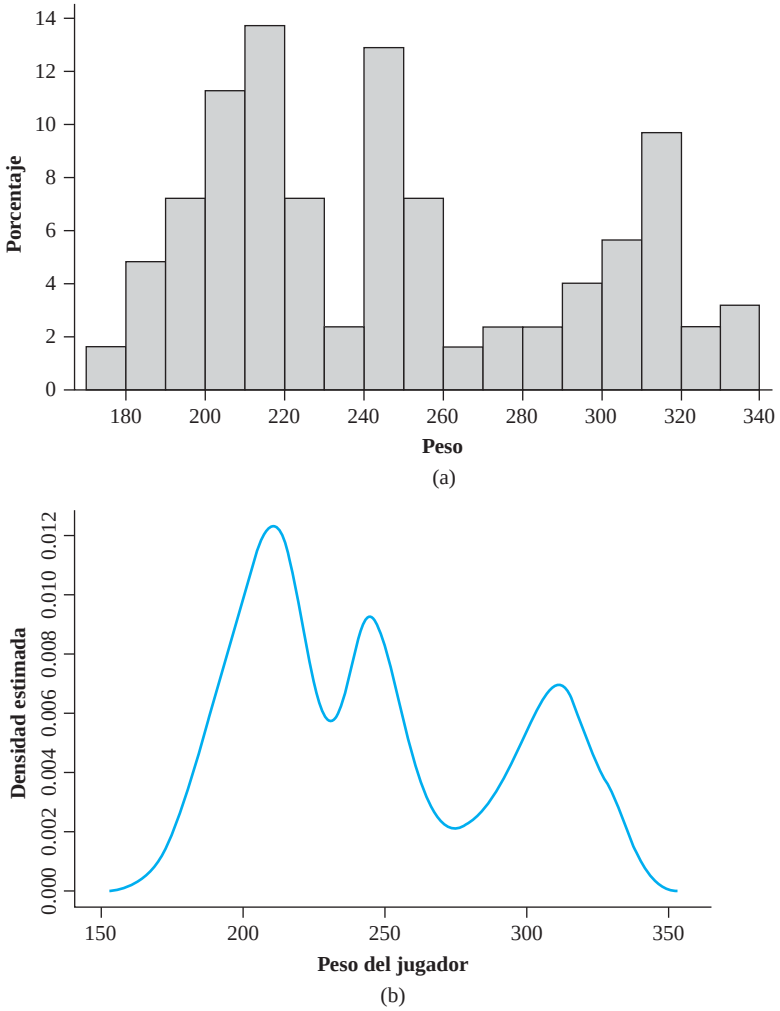


Figura 1.11 Peso de los jugadores de la NFL (a) Histograma (b) Histograma suavizado ■

Un histograma es **simétrico** si la mitad izquierda es una imagen de espejo de la mitad derecha. Un histograma unimodal es **positivamente asimétrico** si la cola derecha o superior se alarga en comparación con la cola izquierda o inferior y **negativamente asimétrico** si el alargamiento es hacia la izquierda. La figura 1.12 muestra histogramas "suavizados" obtenidos superponiendo una curva suavizada sobre los rectángulos, que ilustran las varias posibilidades.

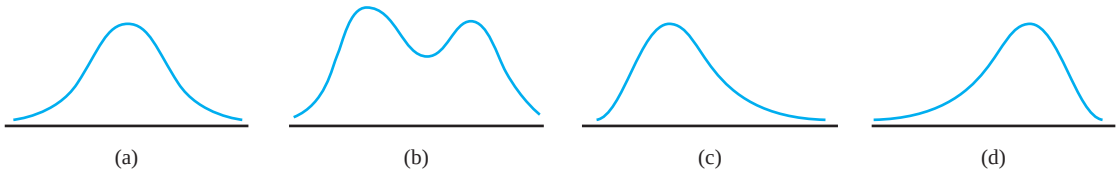


Figura 1.12 Histogramas suavizados: (a) unimodal simétrico; (b) bimodal; (c) positivamente asimétrico y (d) negativamente asimétrico

Datos cualitativos

Tanto una distribución de frecuencia como un histograma pueden ser construidos cuando el conjunto de datos es de naturaleza *cualitativa* (categórico). En algunos casos, habrá un ordenamiento natural de las clases, por ejemplo, estudiantes de primer año, segundo, tercero, cuarto y graduados, mientras que en otros casos el orden será arbitrario, por ejemplo, católico, judío, protestante, etc. Con esos datos categóricos, los intervalos sobre los que se construyen los rectángulos deberán ser de igual ancho.

Ejemplo 1.13 El Public Policy Institute of California realizó una encuesta telefónica de 2501 residentes adultos en California durante abril de 2006 para indagar qué pensaban sobre varios aspectos de la educación pública K-12. Una pregunta fue "en general, ¿cómo calificaría la calidad de las escuelas públicas de su vecindario hoy en día?" La tabla 1.2 muestra las frecuencias y las frecuencias relativas y la figura 1.13 muestra el histograma correspondiente (gráfica de barras).

Tabla 1.2 Distribución de frecuencia para los datos de la calificación de las escuelas

Calificación	Frecuencia	Frecuencia relativa
A	478	.191
B	893	.357
C	680	.272
D	178	.071
F	100	.040
No sabe	172	.069
	2501	1.000

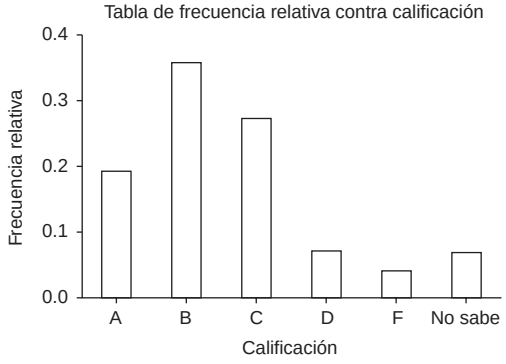


Figura 1.13 Histograma Minitab de los datos de la calificación

Más de la mitad de los encuestados otorgaron una calificación A o B y sólo un poco más de 10% otorgó una calificación D o F. Los porcentajes de padres de niños que asisten a escuelas públicas fueron un poco más favorables para las escuelas: 24%, 40%, 24%, 6%, 4% y 2%. ■

Datos multivariantes

Los datos multivariantes en general son más difíciles de describir en forma visual. Varios métodos para hacerlo aparecen más adelante en el libro, notablemente gráficas de dispersión para datos numéricos bivariantes.

EJERCICIOS Sección 1.2 (10–32)

10. Considere los datos de resistencia de las vigas del ejemplo 1.2.
- **a.** Construya una gráfica de tallos y hojas de los datos. ¿Cuál parece ser el valor de resistencia representativo? ¿Parecen estar las observaciones altamente concentradas en torno al valor representativo o algo dispersas?
- **b.** ¿Parece ser la gráfica razonablemente simétrica en torno a un valor representativo o describiría su forma de otra manera?
- **c.** ¿Parece haber algunos valores de resistencia extremos?
- **d.** ¿Qué proporción de las observaciones de resistencia en esta muestra exceden de 10 MPa?

11. Cada calificación en el siguiente lote de calificaciones de exámenes se encuentra en los 60, 70, 80 o 90. Una gráfica de tallos y hojas con sólo los cuatro tallos 6, 7, 8 y 9 no describiría detalladamente la distribución de calificaciones. En tales situaciones, es deseable utilizar tallos repetidos. En este caso se repetiría el tallo 6 dos veces, utilizando 6B para las calificaciones en los 60 bajos (hojas 0, 1, 2, 3 y 4) y 6A para las calificaciones en los 60 altos (hojas 5, 6, 7, 8 y 9). Asimismo, los demás tallos pueden ser repetidos dos veces para obtener una gráfica de ocho filas. Construya la gráfica para las calificaciones dadas. ¿Qué característica de los datos es resaltada por esta gráfica?

74 89 80 93 64 67 72 70 66 85 89 81 81
71 74 82 85 63 72 81 81 95 84 81 80 70
69 66 60 83 85 98 84 68 90 82 69 72 87
88

12. Los valores de gravedad específica anexos de varios tipos de madera utilizados en la construcción aparecieron en el artículo "Bolted Connection Design Values Based on European Yield Model" (*J. of Structural Engr*., 1993: 2169–2186):

.31 .35 .36 .36 .37 .38 .40 .40 .40
.41 .41 .42 .42 .42 .42 .42 .43 .44
.45 .46 .46 .47 .48 .48 .48 .51 .54
.54 .55 .58 .62 .66 .66 .67 .68 .75

Construya una gráfica de tallos y hojas con tallos repetidos (véase el ejercicio previo) y comente sobre cualquier característica interesante de la gráfica.

13. Las propiedades mecánicas permisibles para el diseño estructural de vehículos aeroespaciales metálicos requieren un método aprobado para analizar estadísticamente datos de prueba empíricos. El artículo "Establishing Mechanical Property Allowables for Metals" (*J. of Testing and Evaluation*, 1998: 293–299) utilizó los datos anexos sobre resistencia a la tensión última (kg/pulg^2) como base para abordar las dificultades que se presentan en el desarrollo de dicho método.

122.2 124.2 124.3 125.6 126.3 126.5 126.5 127.2 127.3
127.5 127.9 128.6 128.8 129.0 129.2 129.4 129.6 130.2
130.4 130.8 131.3 131.4 131.4 131.5 131.6 131.6 131.8
131.8 132.3 132.4 132.4 132.5 132.5 132.5 132.5 132.6
132.7 132.9 133.0 133.1 133.1 133.1 133.1 133.2 133.2
133.2 133.3 133.3 133.5 133.5 133.5 133.8 133.9 134.0
134.0 134.0 134.0 134.1 134.2 134.3 134.4 134.4 134.6
134.7 134.7 134.7 134.8 134.8 134.8 134.9 134.9 135.2
135.2 135.2 135.3 135.3 135.4 135.5 135.5 135.6 135.6
135.7 135.8 135.8 135.8 135.8 135.8 135.9 135.9 135.9
135.9 136.0 136.0 136.1 136.2 136.2 136.3 136.4 136.4
136.6 136.8 136.9 136.9 137.0 137.1 137.2 137.6 137.6
137.8 137.8 137.8 137.9 137.9 138.2 138.2 138.3 138.3
138.4 138.4 138.4 138.5 138.5 138.6 138.7 138.7 139.0
139.1 139.5 139.6 139.8 139.8 140.0 140.0 140.7 140.7
140.9 140.9 141.2 141.4 141.5 141.6 142.9 143.4 143.5
143.6 143.8 143.8 143.9 144.1 144.5 144.5 147.7 147.7

- **a.** Construya una gráfica de tallos y hojas de los datos eliminando (truncando) los dígitos de décimos y luego repitiendo cada valor de tallo cinco veces (una vez para las hojas 1 y 2, una segunda vez para las hojas 3 y 4, etc.). ¿Por qué es relativamente fácil identificar un valor de resistencia representativo?
- **b.** Construya un histograma utilizando clases de ancho igual con la primera clase que tiene un límite inferior de 122 y un límite superior de 124. En seguida comente sobre cualquier característica interesante del histograma.

14. El conjunto de datos adjunto se compone de observaciones del flujo de una regadera (L/min) para una muestra de $n = 129$ casas en Perth, Australia ("An Application of Bayes Methodology to the Analysis of Diary Records in a Water Use Study", *J. Amer. Stat. Assoc*., 1987: 705–711):

4.6 12.3 7.1 7.0 4.0 9.2 6.7 6.9 11.5 5.1
11.2 10.5 14.3 8.0 8.8 6.4 5.1 5.6 9.6 7.5
7.5 6.2 5.8 2.3 3.4 10.4 9.8 6.6 3.7 6.4
8.3 6.5 7.6 9.3 9.2 7.3 5.0 6.3 13.8 6.2
5.4 4.8 7.5 6.0 6.9 10.8 7.5 6.6 5.0 3.3
7.6 3.9 11.9 2.2 15.0 7.2 6.1 15.3 18.9 7.2
5.4 5.5 4.3 9.0 12.7 11.3 7.4 5.0 3.5 8.2
8.4 7.3 10.3 11.9 6.0 5.6 9.5 9.3 10.4 9.7
5.1 6.7 10.2 6.2 8.4 7.0 4.8 5.6 10.5 14.6
10.8 15.5 7.5 6.4 3.4 5.5 6.6 5.9 15.0 9.6
7.8 7.0 6.9 4.1 3.6 11.9 3.7 5.7 6.8 11.3
9.3 9.6 10.4 9.3 6.9 9.8 9.1 10.6 4.5 6.2
8.3 3.2 4.9 5.0 6.0 8.2 6.3 3.8 6.0

- **a.** Construya una gráfica de tallos y hojas de los datos.
- **b.** ¿Cuál es una velocidad de flujo o gasto típico o representativo?
- **c.** ¿Parece estar la gráfica altamente concentrada o dispersa?
- **d.** ¿Es la distribución de valores razonablemente simétrica? Si no, ¿cómo describiría el alejamiento de la simetría?
- **e.** ¿Describiría alguna observación como alejada del resto de los datos (un valor extremo)?

15. ¿Los tiempos de duración de las películas estadounidenses difieren de alguna manera de las del cine francés? El autor investigó esta cuestión seleccionando aleatoriamente 25 películas recientes de cada tipo, lo que resulta en los siguientes tiempos de duración (min):

Am: 94 90 95 93 128 95 125 91 104 116 162 102 90
110 92 113 116 90 97 103 95 120 109 91 138
Fr: 123 116 90 158 122 119 125 90 96 94 137 102 105
106 95 125 122 103 96 111 81 113 128 93 92

Construya una gráfica de tallos y hojas *comparativa* y ponga una lista de tallos a la mitad de la página y luego coloque las hojas Am a la izquierda y las Fr a la derecha. A continuación comente las características interesantes de la gráfica.

16. El artículo citado en el ejemplo 1.2 también dio las observaciones de resistencia adjuntas para los cilindros:

6.1 5.8 7.8 7.1 7.2 9.2 6.6 8.3 7.0 8.3
7.8 8.1 7.4 8.5 8.9 9.8 9.7 14.1 12.6 11.2

a. Construya una gráfica de tallos y hojas comparativa (véase el ejercicio previo) de los datos de la viga y el cilindro y luego responda las preguntas en los incisos (b)–(d) del ejercicio 10 para las observaciones de los cilindros.
b. ¿En qué formas son similares los dos lados de la gráfica? ¿Existen algunas diferencias obvias entre las observaciones de la viga y las observaciones del cilindro?
c. Construya una gráfica de puntos de los datos del cilindro.

17. Transductores de temperatura de cierto tipo se envían en lotes de 50. Se seleccionó una muestra de 60 lotes y se determinó el número de transductores en cada lote que no cumplen con las especificaciones de diseño y se obtuvieron los datos siguientes:

2 1 2 4 0 1 3 2 0 5 3 3 1 3 2 4 7 0 2 3
0 4 2 1 3 1 1 3 4 1 2 3 2 2 8 4 5 1 3 1
5 0 2 3 2 1 0 6 4 2 1 6 0 3 3 3 6 1 2 3

a. Determine las frecuencias y las frecuencias relativas de los valores observados de x = número de transductores en un lote que no cumplen con las especificaciones.
b. ¿Qué proporción de lotes muestreados tienen a lo sumo cinco transductores que no cumplen con las especificaciones? ¿Qué proporción tienen menos de cinco? ¿Qué proporción tienen por lo menos cinco unidades que no cumplen con las especificaciones?
c. Trace un histograma de los datos con la frecuencia relativa en la escala vertical y comente sus características.

18. En un estudio de productividad de autores ("Lotka's Test", *Collection Mgmt*., 1982: 111–118), se clasificó a un gran número de autores de acuerdo con el número de artículos que publicaron durante cierto periodo. Los resultados se presentaron en la distribución de frecuencia adjunta:

Número de artículos	1	2	3	4	5	6	7	8
Frecuencia	784	204	127	50	33	28	19	19

Número de artículos	9	10	11	12	13	14	15	16	17
Frecuencia	6	7	6	7	4	4	5	3	3

a. Construya un histograma correspondiente a esta distribución de frecuencia. ¿Cuál es la característica más interesante de la forma de la distribución?
b. ¿Qué proporción de estos autores publicó por lo menos cinco artículos? ¿Por lo menos diez artículos? ¿Más de diez artículos?
c. Suponga que los cinco 15, los tres 16 y los tres 17 se agruparon en una sola categoría mostrada como "≥15". ¿Podría trazar un histograma? Explique.
d. Suponga que en lugar de que los valores 15, 16 y 17 se enlisten por separado éstos se combinan en la categoría 15–17 con frecuencia 11. ¿Sería capaz de trazar un histograma? Explique.

19. Se determinó el número de partículas contaminante en una oblea de silicio antes de cierto proceso de enjuague para cada oblea en una muestra de tamaño 100 y se obtuvieron las siguientes frecuencias:

Número de partículas	0	1	2	3	4	5	6	7
Frecuencia	1	2	3	12	11	15	18	10
Número de partículas	8	9	10	11	12	13	14	
Frecuencia	12	4	5	3	1	2	1	

a. ¿Qué proporción de las obleas muestreadas tuvieron por lo menos una partícula? ¿Por lo menos cinco partículas?
b. ¿Qué proporción de las obleas muestreadas tuvieron entre cinco y diez partículas, inclusive? ¿Estrictamente entre cinco y diez partículas?
c. Trace un histograma con la frecuencia relativa en el eje vertical. ¿Cómo describiría la forma del histograma?

20. El artículo "Determination of Most Representative Subdivision" (*J. of Energy Engr*., 1993: 43–55) dio datos sobre varias características de subdivisiones que podrían ser utilizadas para decidir si se suministra energía eléctrica con líneas elevadas o líneas subterráneas. He aquí los valores de la variable x = longitud total de calles dentro de una subdivisión:

1280 5320 4390 2100 1240 3060 4770
1050 360 3330 3380 340 1000 960
1320 530 3350 540 3870 1250 2400
960 1120 2120 450 2250 2320 2400
3150 5700 5220 500 1850 2460 5850
2700 2730 1670 100 5770 3150 1890
510 240 396 1419 2109

a. Construya una gráfica de hojas y tallos con el dígito de los millares como tallo y el dígito de las centenas como las hojas y comente sobre las varias características de la gráfica.
b. Construya un histograma con los límites de clase, 0, 1000, 2000, 3000, 4000, 5000 y 6000. ¿Qué proporción de subdivisiones tienen una longitud total menor que 2000? ¿Entre 2000 y 4000? ¿Cómo describiría la forma del histograma?

21. El artículo citado en el ejercicio 20 también da los siguientes valores de las variables y = número de calles cerradas y z = número de intersecciones:

y 1 0 1 0 0 2 0 1 1 1 2 1 0 0 1 1 0 1 1
z 1 8 6 1 1 5 3 0 0 4 4 0 0 1 2 1 4 0 4
y 1 1 0 0 0 1 1 2 0 1 2 2 1 1 0 2 1 1 0
z 0 3 0 1 1 0 1 3 2 4 6 6 0 1 1 8 3 3 5
y 1 5 0 3 0 1 1 0 0
z 0 5 2 3 1 0 0 0 3

a. Construya un histograma con los datos *y*. ¿Qué proporción de estas subdivisiones no tenía calles cerradas? ¿Por lo menos una calle cerrada?

b. Construya un histograma con los datos *z*. ¿Qué proporción de estas subdivisiones tenía cuando mucho cinco intersecciones? ¿Menos de cinco intersecciones?

22. ¿Cómo varía la velocidad de un corredor sobre el curso de un maratón (una distancia de 42.195 km)? Considere determinar tanto el tiempo de recorrido de los primeros 5 km y el tiempo de recorrido entre los 35 y 40 km, y luego reste el primer tiempo del segundo. Un valor positivo de esta diferencia corresponde a un corredor que corre más lento hacia el final de la carrera. El histograma adjunto está basado en tiempos de corredores que participaron en varios maratones japoneses ("Factors Affecting Runners' Maraton Performance", *Chance*, otoño de 1993: 24–30).

¿Cuáles son algunas características interesantes de este histograma? ¿Cuál es un valor de diferencia típico? ¿Aproximadamente qué proporción de los corredores corren la última distancia más rápido que la primera?

23. El artículo "Statistical Modeling of the Time Course of Tantrum Anger" (*Annals of Applied Stats*, 2009: 1013–1034) analizó cómo la intensidad de la ira en los berrinches de los niños podría estar relacionada con la duración de la rabieta, así como los indicadores de comportamiento, tales como gritar, arañar y empujar o tirar. Se proporcionó la distribución de frecuencias siguiente (y también el histograma correspondiente):

0−<2 :	136	2−<4:	92	4−<11:	71
11−<20:	26	20−<30:	7	30−<40:	3

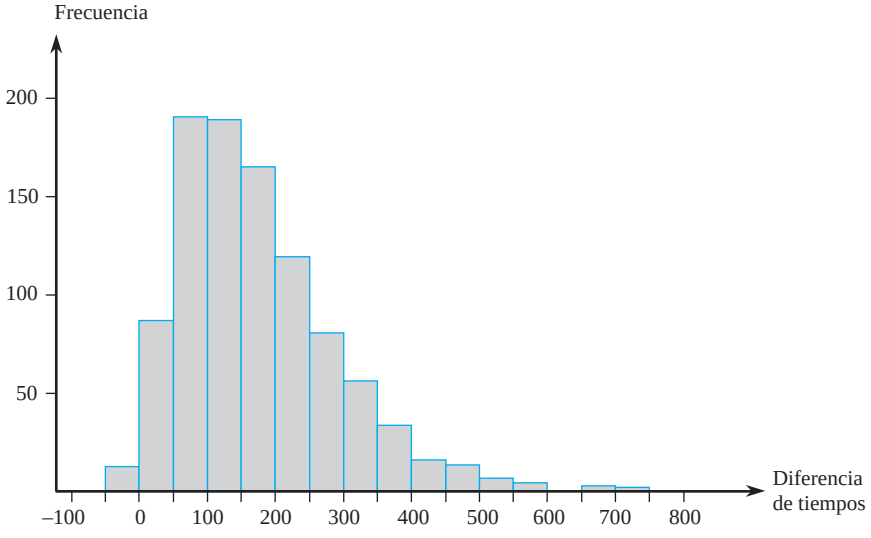
Construya un histograma y comente sobre las características interesantes.

24. El conjunto de datos adjuntos consiste en observaciones de resistencia al esfuerzo cortante (lb) de soldaduras de puntos ultrasónicas aplicadas en un cierto tipo de lámina alclad. Construya un histograma de frecuencia relativa basado en diez clases de ancho igual con límites 4000, 4200, ... [El histograma concordará con el que aparece en "Comparison of Properties of Joints Prepared by Ultrasonic Welding and Other Means" (*J. of Aircraft*, 1983: 552–556).] Comente sobre sus características.

5434	4948	4521	4570	4990	5702	5241
5112	5015	4659	4806	4637	5670	4381
4820	5043	4886	4599	5288	5299	4848
5378	5260	5055	5828	5218	4859	4780
5027	5008	4609	4772	5133	5095	4618
4848	5089	5518	5333	5164	5342	5069
4755	4925	5001	4803	4951	5679	5256
5207	5621	4918	5138	4786	4500	5461
5049	4974	4592	4173	5296	4965	5170
4740	5173	4568	5653	5078	4900	4968
5248	5245	4723	5275	5419	5205	4452
5227	5555	5388	5498	4681	5076	4774
4931	4493	5309	5582	4308	4823	4417
5364	5640	5069	5188	5764	5273	5042
5189	4986					

25. Una transformación de valores de datos por medio de alguna función matemática, tal como $\sqrt{x}$ o $1/x$ a menudo produce un conjunto de números que tienen "mejores" propiedades estadísticas que los datos originales. En particular, puede ser posible encontrar una función para la cual el histograma de valores transformados es más simétrico (o, incluso, mejor, más como una curva en forma de campana) que los datos originales. Por ejemplo, el artículo "Time Lapse Cinematographic Analysis of Beryllium-Lung Fibroblast Interactions" (*Environ. Research*, 1983: 34–43) reportó los resultados de experimentos diseñados para estudiar el comportamiento de ciertas células individuales que habían estado expuestas a berilio. Una importante característica de dichas células individuales es su tiempo de interdivisión (IDT, por sus siglas en inglés). Se determinaron tiempos de interdivisión de un gran número de células, tanto en condiciones expuestas (tratamiento) como no expuestas (control). Los autores del artículo utilizaron una transformación logarítmica,

Histograma para el ejercicio 22



es decir, valor transformado = log(valor original). Considere los siguientes tiempos de interdivisión representativos.

IDT	$\log_{10}$(IDT)	IDT	$\log_{10}$(IDT)	IDT	$\log_{10}$(IDT)
28.1	1.45	60.1	1.78	21.0	1.32
31.2	1.49	23.7	1.37	22.3	1.35
13.7	1.14	18.6	1.27	15.5	1.19
46.0	1.66	21.4	1.33	36.3	1.56
25.8	1.41	26.6	1.42	19.1	1.28
16.8	1.23	26.2	1.42	38.4	1.58
34.8	1.54	32.0	1.51	72.8	1.86
62.3	1.79	43.5	1.64	48.9	1.69
28.0	1.45	17.4	1.24	21.4	1.33
17.9	1.25	38.8	1.59	20.7	1.32
19.5	1.29	30.6	1.49	57.3	1.76
21.1	1.32	55.6	1.75	40.9	1.61
31.9	1.50	25.5	1.41		
28.9	1.46	52.1	1.72		

Use los intervalos de clase 10−<20, 20−<30, . . . para construir un histograma de los datos originales. Use los intervalos 1.1−<1.2, 1.2−<1.3, . . . para hacer lo mismo con los datos transformados. ¿Cuál es el efecto de la transformación?

26. En la actualidad se está utilizando la difracción retrodispersada de electrones en el estudio de fenómenos de fractura. La siguiente información sobre ángulo de desorientación (grados) se extrajo del artículo "Observations on the Faceted Initiation Site in the Dwell-Fatigue Tested Ti-6242 Alloy: Crystallographic Orientation and Size Effects" (*Metallurgical and Materials Trans*., 2006: 1507–1518).

Clase:	0−<5	5−<10	10−<15	15−<20
Frecuencia relativa:	.177	.166	.175	.136
Clase:	20−<30	30−<40	40−<60	60−<90
Frecuencia relativa:	.194	.078	.044	.030

a. ¿Es verdad que más del 50% de los ángulos muestreados son más pequeños que 15°, como se afirma en el artículo?
b. ¿Qué proporción de los ángulos muestreados son por lo menos de 30°?
c. ¿Aproximadamente qué proporción de los ángulos son de entre 10° y 25°?
d. Construya un histograma y comente sobre cualquier característica interesante.

27. El artículo "Study on the Life Distribution of Microdrills" (*J. of Engr. Manufacture*, 2002: (301–305) reportó las siguientes observaciones, listadas en orden creciente sobre la duración de brocas (número de agujeros que una broca fresa antes de que se rompa) cuando se fresaron agujeros en una cierta aleación de latón.

11	14	20	23	31	36	39	44	47	50
59	61	65	67	68	71	74	76	78	79
81	84	85	89	91	93	96	99	101	104
105	105	112	118	123	136	139	141	148	158
161	168	184	206	248	263	289	322	388	513

a. ¿Por qué una distribución de frecuencia no puede estar basada en los intervalos de clase 0–50, 50–100, 100–150 y así sucesivamente?
b. Construya una distribución de frecuencia e histograma de los datos con los límites de clase 0, 50, 100, . . . , y luego comente sobre las características interesantes.
c. Construya una distribución de frecuencia e histograma de los logaritmos naturales de las observaciones de duración y comente sobre las características interesantes.
d. ¿Qué proporción de las observaciones de duración en esta muestra son menores que 100? ¿Qué proporción de las observaciones son de por lo menos 200?

28. Las mediciones humanas constituyen una rica área de aplicación de métodos estadísticos. El artículo "A Longitudinal Study of the Development of Elementary School Children's Private Speech" (*Merrill-Palmer Q*., 1990: 443–463) reportó sobre un estudio de niños que hablan solos (conversación a solas). Se pensaba que la convesación a solas tenía que ver con el IQ, porque se supone que éste mide la madurez mental y se sabía que la convesación a solas disminuye conforme los estudiantes avanzan a través de los años de la escuela primaria. El estudio incluyó 33 estudiantes cuyas calificaciones de IQ de primer año se dan a continuación:

82	96	99	102	103	103	106	107	108	108	108	108	
109	110	110	111	113	113	113	113	115	115	118	118	
119	121	122	122	127	132	136	140	146				

Describa los datos y comente sobre cualquier característica importante.

29. Considere los siguientes datos sobre el tipo de problemas de salud (J = hinchazón de las articulaciones, F = fatiga, B = dolor de espalda, M = debilidad muscular, C = tos, N = escurrimiento nasal/irritación, O = otro) que aquejan a los plantadores de árboles. Obtenga las frecuencias y las frecuencias relativas de las diversas categorías y trace un histograma. (Los datos son consistentes con los porcentajes dados en el artículo "Physiological Effects of Work Stress and Pesticide Exposure in Tree Planting by British Columbia Silviculture Workers", *Ergonomics*, 1993: 951–961.)

O	O	N	J	C	F	B	B	F	O	J	O	O	M
O	F	F	O	O	N	O	N	J	F	J	B	O	C
J	O	J	J	F	N	O	B	M	O	J	M	O	B
O	F	J	O	O	B	N	C	O	O	O	M	B	F
J	O	F	N										

30. Un **diagrama de Pareto** es una variación de un histograma de datos categóricos producidos por un estudio de control de calidad. Cada categoría representa un tipo diferente de no conformidad del producto o problema de producción. Las categorías se ordenaron de modo que la categoría con la frecuencia más grande aparezca a la extrema izquierda, luego la categoría con la segunda frecuencia más grande, y así sucesivamente. Suponga que se obtiene la siguiente información sobre no conformidades en paquetes de circuito: componentes averiados, 126; componentes incorrectos, 210; soldadura insuficiente, 67; soldadura excesiva, 54; componente faltante, 131. Construya un diagrama de Pareto.

31. La **frecuencia acumulativa** y la frecuencia relativa acumulativa de un intervalo de clase particular son la suma de frecuencias y frecuencias relativas, respectivamente, del intervalo y todos los intervalos que quedan debajo de él. Si, por ejemplo,

existen cuatro intervalos con frecuencias 9, 16, 13 y 12, entonces las frecuencias acumulativas son 9, 25, 38 y 50 y las frecuencias relativas acumulativas son .18, .50, .76 y 1.00. Calcule las frecuencias acumulativas y las frecuencias relativas acumulativas de los datos del ejercicio 24.

32. La carga de fuego (MJ/m^2) es la energía calorífica que podría ser liberada por metro cuadrado de área de piso por la combustión del contenido y la estructura misma. El artículo "Fire Loads in Office Buildings" (*J. of Structural Engr.*, 1997: 365–368) dio los siguientes porcentajes acumulativos (tomados de una gráfica) de cargas de fuego en una muestra de 388 cuartos:

Valor	0	150	300	450	600
% acumulativo	0	19.3	37.6	62.7	77.5
Valor	750	900	1050	1200	1350
% acumulativo	87.2	93.8	95.7	98.6	99.1
Valor	1500	1650	1800	1950	
% acumulativo	99.5	99.6	99.8	100.0	

a. Construya un histograma de frecuencia relativa y comente sobre las características interesantes.
b. ¿Qué proporción de cargas de fuego son menores que 600? ¿Por lo menos de 1200?
c. ¿Qué proporción de las cargas está entre 600 y 1200?

1.3 Medidas de ubicación

Los resúmenes visuales de datos son herramientas excelentes para obtener impresiones y percepciones preliminares. Un análisis de datos más formal a menudo requiere el cálculo e interpretación de medidas resumidas numéricas. Es decir, de los datos se trata de extraer varios números resumidos, números que podrían servir para caracterizar el conjunto de datos y comunicar algunas de sus características prominentes. El interés principal se concentrará en los datos numéricos; al final de la sección aparecen algunos comentarios con respecto a datos categóricos.

Supóngase, entonces, que el conjunto de datos es de la forma $x_1, x_2, \ldots, x_n$, donde cada x_i es un número. ¿Qué características del conjunto de números son de mayor interés y merecen énfasis? Una importante característica de un conjunto de números es su ubicación y en particular su centro. Esta sección presenta métodos para describir la ubicación de un conjunto de datos; en la sección 1.4 se regresará a los métodos para medir la variabilidad en un conjunto de números.

La media

Para un conjunto dado de números $x_1, x_2, \ldots, x_n$, la medida más conocida y útil del centro es la *media* o promedio aritmético del conjunto. Como casi siempre se pensará que los números x_i constituyen una muestra, a menudo se hará referencia al promedio aritmético como la *media muestral* y se la denotará por $\bar{x}$.

DEFINICIÓN

La **media muestral** $\bar{x}$ de las observaciones $x_1, x_2, \ldots, x_n$ está dada por

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

El numerador de $\bar{x}$ se escribe más informalmente como $\sum x_i$, donde la suma incluye todas las observaciones muestrales.

Para reportar $\bar{x}$, se recomienda utilizar una precisión decimal de un dígito más que la precisión de los números x_i. Por consiguiente si las observaciones son distancias de detención con $x_1 = 125$, $x_2 = 131$, y así sucesivamente, se podría tener $\bar{x} = 127.3$ pies.

Ejemplo 1.14 El agrietamiento de hierro y acero provocado por corrosión producida por esfuerzo cáustico ha sido estudiado debido a las fallas que se presentan alrededor de los remaches en calderas de acero y fallas de rotores de turbinas de vapor. Considérense las observaciones adjuntas de x = longitud de agrietamiento (μm) derivadas de pruebas de corrosión con esfuerzo constante en probetas de barras pulidas sometidas a tensión durante un lapso de tiempo fijo. (Los datos concuerdan con un histograma y cantidades resumidas tomadas del artículo "On the Role of Phosphorus in the Caustic Stress Corrosion Cracking of Low Alloy Steels", *Corrosion Science*, 1989: 53–68.)

$$\begin{aligned} x_1 &= 16.1 \; x_2 = 9.6 \; x_3 = 24.9 \; x_4 = 20.4 \; x_5 = 12.7 \; x_6 = 21.2 \; x_7 = 30.2 \\ x_8 &= 25.8 \; x_9 = 18.5 \; x_{10} = 10.3 \; x_{11} = 25.3 \; x_{12} = 14.0 \; x_{13} = 27.1 \; x_{14} = 45.0 \\ x_{15} &= 23.3 \; x_{16} = 24.2 \; x_{17} = 14.6 \; x_{18} = 8.9 \; x_{19} = 32.4 \; x_{20} = 11.8 \; x_{21} = 28.5 \end{aligned}$$

La figura 1.14 muestra una gráfica de tallo y hojas de los datos; una longitud de agrietamiento en los 20 bajos parece ser "típica".

0A	96	89			
1B	27	03	40	46	18
1A	61	85			
2B	49	04	12	33	42
2A	58	53	71	85	
3B	02	24			
3A					
4B					
4A	50				

Tallo: dígito de las decenas
Hoja: dígito de las unidades y décimas

Figura 1.14 Gráfica de tallo y hojas de los datos de la longitud de agrietamiento

Con $\sum x_i = 444.8$, la media muestral es

$$\bar{x} = \frac{444.8}{21} = 21.18$$

un valor consistente con la información dada por la gráfica de tallo y hojas. ■

Una interpretación física de $\bar{x}$ demuestra cómo mide la ubicación (centro) de una muestra. Se traza y gradúa un eje de medición horizontal y luego se representa cada observación muestral por una pesa de 1 lb colocada en el punto correspondiente sobre el eje. El único punto en el cual se puede colocar un punto de apoyo para equilibrar el sistema de pesos es el punto correspondiente al valor de $\bar{x}$ (véase la figura 1.15).

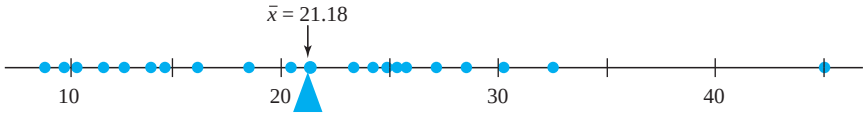


Figura 1.15 La media es el punto de equilibrio para un sistema de pesos

Así como $\bar{x}$ representa el valor promedio de las observaciones incluidas en una muestra, se puede calcular el promedio de todos los valores incluidos en la población. Este promedio se llama **media de la población** y está denotado por la letra griega μ. Cuando existen N valores en la población (una población finita), entonces μ = (suma de los N valores de población)/N. En los capítulos 3 y 4, se dará una definición más general de μ que se aplica tanto a poblaciones finitas y (conceptualmente) infinitas. Así como $\bar{x}$ es una medida interesante e importante de la ubicación de la muestra, μ es una interesante e importante característica (con frecuencia la más importante) de una población. En los capí-

tulos sobre inferencia estadística, se presentarán métodos basados en la media muestral para sacar conclusiones con respecto a una media de población. Por ejemplo, se podría utilizar la media muestral $\bar{x} = 21.18$ calculada en el ejemplo 1.14 como una *estimación puntual* (un solo número que es la "mejor" conjetura) de μ = la longitud de agrietamiento promedio verdadera de todas las probetas tratadas como se describe.

La media sufre de una deficiencia que la hace ser una medida inapropiada del centro en algunas circunstancias: su valor puede ser afectado en gran medida por la presencia de incluso un solo valor extremo (una observación inusualmente grande o pequeña). En el ejemplo 1.14, el valor $x_{14} = 45.0$ es obviamente un valor extremo. Sin esta observación, $\bar{x} = 399.8/20 = 19.99$; el valor extremo incrementa la media en más de 1 μm. Si la observación de 45.0 μm fuera reemplazada por el valor catastrófico de 295.0 μm, un valor realmente extremo, entonces $\bar{x} = 694.8/21 = 33.09$, ¡el cual es más grande que todas excepto una de las observaciones!

Una muestra de ingresos a menudo produce algunos valores apartados (unos cuantos afortunados que ganan cantidades astronómicas) y el uso del ingreso promedio como medida de ubicación con frecuencia será engañoso. Tales ejemplos sugieren que se busca una medida que sea menos sensible a los valores apartados que $\bar{x}$ y momentáneamente se propondrá una. Sin embargo, aunque $\bar{x}$ sí tiene este defecto potencial, sigue siendo la medida más ampliamente utilizada, en gran medida porque existen muchas poblaciones para las cuales un valor extremo en la muestra sería altamente improbable. Cuando se muestrea una población como ésa (una población normal o en forma de campana es el ejemplo más importante), la media muestral tenderá a ser estable y bastante representativa de la muestra.

La mediana

La palabra *mediana* es sinónimo de "medio" y la media muestral es en realidad el valor medio una vez que se ordenan las observaciones de la más pequeña a la más grande. Cuando las observaciones están denotadas por $x_1, \ldots, x_n$, se utilizará el símbolo $\tilde{x}$ para representar la mediana muestral.

DEFINICIÓN

La **mediana muestral** se obtiene ordenando primero las n observaciones de la más pequeña a la más grande (con cualesquiera valores repetidos incluidos de modo que cada observación muestral aparezca en la lista ordenada). Entonces,

$$\tilde{x} = \begin{cases} \text{El valor medio único si } n \text{ es impar.} & = \left(\dfrac{n+1}{2}\right)^{\text{ésimo}} \text{ valor ordenado} \\ \text{El promedio de los dos valores medios si } n \text{ es par;} & = \text{promedio de } \left(\dfrac{n}{2}\right)^{\text{ésimo}} \text{ y } \left(\dfrac{n}{2}+1\right)^{\text{ésimo}} \text{ valores ordenados} \end{cases}$$

Ejemplo 1.15 Las personas que no están familiarizadas con la música clásica pueden tender a creer que las instrucciones de un compositor para la reproducción de una pieza en particular son tan específicas que la duración no depende en absoluto del(los) intérprete(s). Sin embargo, normalmente hay un montón de espacio para la interpretación y para que los directores de orquesta y músicos puedan sacar el máximo provecho de ello. El autor se dirigió al sitio

Web ArkivMusic.com y seleccionó una muestra de 12 grabaciones de la Sinfonía # 9 de Beethoven (el "Coral", una obra impresionante y hermosa), generando las duraciones siguientes (en minutos) clasificadas en orden creciente:

62.3 62.8 63.6 65.2 65.7 66.4 67.4 68.4 68.8 70.8 75.7 79.0

He aquí una gráfica de puntos de los datos:

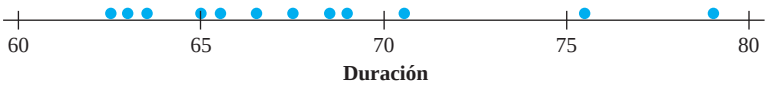


Figura 1.16 Gráfica de puntos de los datos para el ejemplo 1.14

Puesto que $n = 12$ es par, la mediana de la muestra es el promedio de los $n/2 = 6^{o}$ y $(n/2 + 1) = 7^{o}$ valores de la lista ordenada:

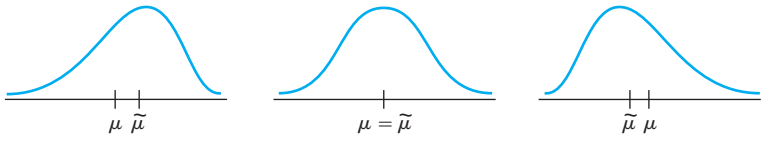
$$\tilde{x} = \frac{66.4 + 67.4}{2} = 66.90$$

Note que si la observación más grande, 79.0, no hubiera aparecido en la muestra, la mediana muestral resultante de las $n = 11$ observaciones habría sido el valor medio 66.4 (el $[n + 1]/2 = 6^{o}$ valor ordenado, es decir el sexto valor contado desde cualquier extremo de la lista ordenada). La media muestral es $\bar{x} = \sum x_i = 816.1/12 = 68.01$, la cual es un poco más de un minuto más grande que la mediana. La media se sale un poco con respecto a la mediana ya que la muestra "se extiende" un poco más en el extremo superior que en el extremo inferior. ■

Los datos del ejemplo 1.15 ilustran una importante propiedad de $\tilde{x}$ en contraste con $\bar{x}$. La mediana muestral es muy insensible a los valores apartados. Si, por ejemplo, las dos x_i más grandes se incrementan desde 75.7 y 79.0 hasta 85.7 y 89.0, respectivamente, $\tilde{x}$ no se vería afectada. Por lo tanto, en el tratamiento de valores apartados, $\bar{x}$ y $\tilde{x}$ no son extremos opuestos de un espectro. Ambas cantidades describen el lugar donde se centran los datos, pero en general no serán iguales porque se enfocan en aspectos diferentes de la muestra.

Análogo a $\tilde{x}$ como valor medio de la muestra existe un valor medio de la población, la **mediana poblacional,** denotada por $\tilde{\mu}$. Como con $\bar{x}$ y μ, se puede pensar en utilizar la mediana muestral $\tilde{x}$ para hacer una inferencia sobre $\tilde{\mu}$. En el ejemplo 1.15, se podría utilizar $\tilde{x} = 66.90$ como estimación de la mediana de tiempo para la población de todas las grabaciones. A menudo se utiliza una mediana para describir ingresos o salarios (debido a que no es influida en gran medida por unos pocos salarios grandes). Si el salario mediano de una muestra de ingenieros fuera $\tilde{x} = \$66{,}416$ dólares éste se podría utilizar como base para concluir que el salario mediano de todos los ingenieros es de más de 60,000 dólares.

La media μ y la mediana $\tilde{\mu}$ poblacionales en general no serán idénticas. Si la distribución de la población es positiva o negativamente asimétrica, como se ilustra en la figura 1.17, entonces $\mu \neq \tilde{\mu}$. Cuando éste es el caso, al hacer inferencias primero se debe decidir cuál de las dos características de la población es de mayor interés y luego proceder como corresponda.



(a) Sesgo negativo (b) Simétrica (c) Sesgo positivo

Figura 1.17 Tres formas diferentes de distribución de la población

Otras medidas de ubicación: cuartiles, percentiles y medias recortadas

La mediana (poblacional o muestral) divide el conjunto de datos en dos partes iguales. Para obtener medidas de ubicación más finas, se podrían dividir los datos en más de dos partes. Tentativamente, los cuartiles dividen el conjunto de datos en cuatro partes iguales y las observaciones arriba del tercer cuartil constituyen el cuarto superior del conjunto de datos, el segundo cuartil es idéntico a la mediana y el primer cuartil separa el cuarto inferior de los tres cuartos superiores. Asimismo, un conjunto de datos (muestra o población) puede ser incluso más finamente dividido por medio de percentiles, el 99º percentil separa el 1% más alto del 99% más bajo, y así sucesivamente. A menos que el número de observaciones sea un múltiplo de 100, se debe tener cuidado al obtener percentiles. En el capítulo 4 se utilizarán percentiles en conexión con ciertos modelos de poblaciones infinitas y por tanto su discusión se pospone hasta ese punto.

La media es bastante sensible a un solo valor extremo, mientras que la mediana es insensible a muchos valores apartados. Como el comportamiento extremo de uno u otro tipo podría ser indeseable, se consideran brevemente medidas alternativas que no son ni sensibles como $\bar{x}$ ni tan insensibles como $\tilde{x}$. Para motivar estas alternativas, obsérvese que $\bar{x}$ y $\tilde{x}$ se encuentran en extremos opuestos de la misma "familia" de medidas. La media es el promedio de todos los datos, mientras que la mediana resulta de eliminar todos excepto uno o dos valores medios y luego promediar. Parafraseando, la media implica recortar 0% de cada extremo de la muestra, mientras que en el caso de la mediana se recorta la cantidad máxima posible de cada extremo. Una **media recortada** es un compromiso entre $\bar{x}$ y $\tilde{x}$. Una media 10% recortada, por ejemplo, se calcularía eliminando el 10% más pequeño y el 10% más grande de la muestra y luego promediando lo que queda.

Ejemplo 1.16 La producción de Bidri es una artesanía tradicional de India. Las artesanías Bidri (tazones, recipientes, etc.) se funden con una aleación que contiene principalmente zinc y algo de cobre. Considere las siguientes observaciones sobre contenido de cobre (%) de una muestra de artefactos Bidri tomada del Museo Victoria y Albert de Londres ("Enigmas of Bidri", *Surface Engr*., 2005: 333–339), enlistadas en orden creciente:

2.0 2.4 2.5 2.6 2.6 2.7 2.7 2.8 3.0 3.1 3.2 3.3 3.3
3.4 3.4 3.6 3.6 3.6 3.6 3.7 4.4 4.6 4.7 4.8 5.3 10.1

La figura 1.18 es una gráfica de puntos de los datos. Una característica prominente es el valor extremo único en el extremo superior; la distribución está un tanto más dispersa en la región de valores grandes que en el caso de valores pequeños. La media muestral y la mediana son 3.65 y 3.35, respectivamente. Se obtiene una media recortada con un porcentaje de recorte de $100(2/26) = 7.7\%$ al eliminar las dos observaciones más pequeñas y las dos más grandes; esto da $\bar{x}_{rec(7.7)} = 3.42$. El recorte en este caso elimina el valor extremo más grande y por tanto acerca la media recortada hacia la mediana.

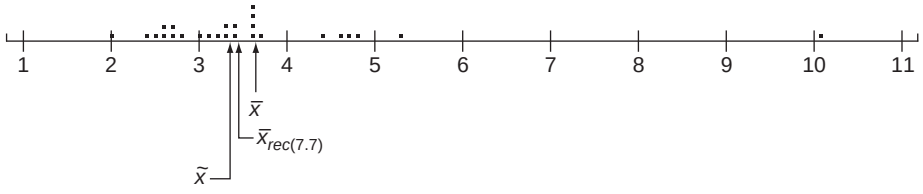


Figura 1.18 Gráfica de puntos del contenido de cobre para el ejemplo 1.16 ■

Una media recortada con un porcentaje de recorte moderado, algo entre 5% y 25%, producirá una medida del centro que no es ni tan sensible a los valores apartados como la media ni tan insensible como la mediana. Si el porcentaje de recorte deseado es $100\alpha\%$ y $n\alpha$ no es un entero, la media recortada debe ser calculada por interpolación. Por ejemplo, considérese $\alpha = .10$ para un porcentaje de recorte de 10% y $n = 26$ como en el ejemplo 1.16. Entonces $\bar{x}_{rec(10)}$ sería el promedio ponderado apropiado de la media recortada 7.7% calculada allí y la media recortada 11.5% que resulta de recortar tres observaciones de cada extremo.

Datos categóricos y proporciones muestrales

Cuando los datos son categóricos, una distribución de frecuencia o una distribución de frecuencia relativa proporciona un resumen tabular efectivo de los datos. Las cantidades resumidas numéricas naturales en esta situación son las frecuencias individuales y las frecuencias relativas. Por ejemplo, si se realiza una encuesta de personas que poseen cámaras digitales para estudiar la preferencia de marcas y cada persona en la muestra identifica la marca de cámara que él o ella posee, entonces se podría contar el número que poseen Canon, Sony, Kodak, y así sucesivamente. Considérese muestrear una población dividida en dos partes, una que consiste en sólo dos categorías (tal como votó o no votó en la última elección, si posee o no una cámara digital, etc.). Si x denota el número en la muestra que cae en la categoría 1, entonces el número en la categoría 2 es $n - x$. La frecuencia relativa o *proporción muestral* en la categoría 1 es x/n y la proporción muestral en la categoría 2 es $1 - x/n$. Designemos con 1 una respuesta que cae en la categoría 1 y con 0 una que cae en la categoría 2. Un tamaño de muestra de $n = 10$ podría dar entonces las respuestas 1, 1, 0, 1, 1, 1, 0, 0, 1, 1. La media muestral de esta muestra numérica es (como la cantidad de números $1 = x = 7$)

$$\frac{x_1 + \cdots + x_n}{n} = \frac{1 + 1 + 0 + \cdots + 1 + 1}{10} = \frac{7}{10} = \frac{x}{n} = \text{proporción muestral}$$

Más generalmente, *enfóquese la atención en una categoría particular y codifíquense los resultados de modo que se anote un* 1 *para una observación comprendida en la categoría y un* 0 *para una observación no comprendida en la categoría. Entonces la proporción muestral de observaciones comprendidas en la categoría es la media muestral de la secuencia de los* 1 y *los* 0. Por consiguiente se puede utilizar una media muestral para resumir los resultados de una muestra categórica. Estos comentarios también se aplican a situaciones en las cuales las categorías se definen agrupando valores en una muestra o población numérica (p. ej., podría existir interés en saber si las personas han tenido su automóvil actual durante por lo menos 5 años, en lugar de estudiar la duración exacta de la tenencia).

Análogo a la proporción muestral x/n de personas u objetos que caen en una categoría particular, represente con p la proporción de aquellos presentes en la población entera que caen en la categoría. Como con x/n, p es una cantidad entre 0 y 1 y mientras que x/n es una característica de la muestra, p es una característica de la población. La relación entre las dos es igual a la relación entre $\tilde{x}$ y $\tilde{\mu}$ y entre $\bar{x}$ y μ. En particular, subsecuentemente se utilizará x/n para hacer inferencias sobre p. Si, por ejemplo, una muestra de 100 propietarios de automóviles reveló que 22 tenían su automóvil desde por lo menos 5 años atrás, en tal caso se podría utilizar $22/100 = .22$ como estimación puntual de la proporción de todos los propietarios que tenían su automóvil desde por lo menos 5 años atrás. Con k categorías ($k > 2$), se pueden utilizar las k proporciones muestrales para responder preguntas sobre las proporciones de población $p_1, \ldots, p_k$.

EJERCICIOS Sección 1.3 (33–43)

33. El 1 de mayo de 2009 *The Montclarian* reportó los siguientes aumentos a los precios de venta de una muestra de casas en Alameda, CA., después de las que se vendieron el mes anterior (miles de dólares):

590 815 575 608 350 1285 408 540 555 679

a. Calcule e interprete la media y la mediana muestrales.
b. Suponga que la 6ª observación hubiera sido 985 en lugar de 1285. ¿Cómo cambiarían las media y la mediana?
c. Calcule una media recortada 20% eliminando primero las dos observaciones muestrales más pequeñas y las dos más grandes.
d. Calcule una media recortada 15%.

34. La exposición a productos microbianos, especialmente endotoxina, puede tener un impacto en la vulnerabilidad a enfermedades alérgicas. El artículo "Dust Sampling Methods for Endotoxin–An Essential, But Underestimated Issue" (*Indoor Air*, 2006: 20–27) consideró temas asociados con la determinación de concentración de endotoxina. Los siguientes datos sobre concentración (EU/mg) en polvo asentado de una muestra de hogares urbanos y otra de casas campestres fueron amablemente suministrados por los autores del artículo citado.

U: 6.0 5.0 11.0 33.0 4.0 5.0 80.0 18.0 35.0 17.0 23.0
C: 4.0 14.0 11.0 9.0 9.0 8.0 4.0 20.0 5.0 8.9 21.0
9.2 3.0 2.0 0.3

a. Determine la media muestral de cada muestra. ¿Cómo se comparan?
b. Determine la mediana muestral de cada muestra. ¿Cómo se comparan? ¿Por qué es la mediana de la muestra urbana tan diferente de la media de dicha muestra?
c. Calcule la media recortada de cada muestra eliminando la observación más pequeña y la más grande. ¿Cuáles son los porcentajes de recorte correspondientes? ¿Cómo se comparan los valores de estas medias recortadas con las medias y medianas correspondientes?

35. La presión de inyección mínima (lb/pulg^2) de especímenes moldeados por inyección de fécula de maíz se determinó con ocho especímenes diferentes (la presión más alta corresponde a una mayor dificultad de procesamiento) y se obtuvieron las siguientes observaciones (tomadas de "Thermoplastic Starch Blends with a Polyethylene-Co-Vinyl Alcohol: Processability and Physical Properties", *Polymer Engr. and Science*, 1994: 17–23):

15.0 13.0 18.0 14.5 12.0 11.0 8.9 8.0

a. Determine los valores de la media muestral, la mediana muestral y la media recortada 12.5% y compare estos valores.
b. ¿En cuánto se podría incrementar la observación más pequeña de la muestra, actualmente 8.0, sin afectar el valor de la mediana muestral?
c. Suponga que desea los valores de la media y la mediana muestrales cuando las observaciones están expresadas en kilogramos por pulgada cuadrada (kg/pulg^2) en lugar de lb/pulg^2. ¿Es necesario volver a expresar cada observación en kg/pulg^2 o se pueden utilizar los valores calculados en el inciso (a) directamente? [*Sugerencia*: 1 kg = 2.2 lb.]

36. Una muestra de 26 trabajadores de plataforma petrolera marina tomaron parte en un ejercicio de escape y se obtuvieron los datos adjuntos de tiempo (s) para completar el escape ("Oxygen Consumption and Ventilation During Escape from an Offshore Platform", *Ergonomics*, 1997: 281–292):

389 356 359 363 375 424 325 394 402
373 373 370 364 366 364 325 339 393
392 369 374 359 356 403 334 397

a. Construya una gráfica de tallo y hojas de los datos. ¿Cómo sugiere la gráfica que la media y mediana muestrales se comparen?
b. Calcule los valores de la media y mediana muestrales [*Sugerencia*: $\sum x_i = 9638$.]
c. ¿En cuánto se podría incrementar el tiempo más largo, actualmente de 424, sin afectar el valor de la mediana muestral? ¿En cuánto se podría disminuir este valor sin afectar el valor de la mediana muestral?
d. ¿Cuáles son los valores de $\bar{x}$ y $\tilde{x}$ cuando las observaciones se reexpresan en minutos?

37. El artículo "Snow Cover and Temperature Relationships in North America and Eurasia" (*J. Climate and Applied Meteorology*, 1983: 460–469) utilizó técnicas estadísticas para relacionar la cantidad de cobertura de nieve sobre cada continente para promediar la temperatura continental. Los datos allí presentados incluyeron las siguientes diez observaciones de la cobertura de nieve en octubre en Eurasia durante los años 1970–1979 (en millones de km^2):

6.5 12.0 14.9 10.0 10.7 7.9 21.9 12.5 14.5 9.2

¿Qué reportaría como valor representativo, o típico, de cobertura de nieve en octubre durante este periodo y qué motivaría su elección?

38. Los valores de presión sanguínea a menudo se reportan a los 5 mmHg más cercanos (100, 105, 110, etc.). Suponga que los valores de presión sanguínea reales de nueve individuos seleccionados al azar son

118.6 127.4 138.4 130.0 113.7 122.0 108.3
131.5 133.2

a. ¿Cuál es la mediana de los valores de presión sanguínea *reportados*?
b. Suponga que la presión sanguínea del segundo individuo es 127.6 en lugar de 127.4 (un pequeño cambio en un solo valor). ¿Cómo afecta esto a la mediana de los valores reportados? ¿Qué dice esto sobre la sensibilidad de la mediana al redondeo o agrupamiento de los datos?

39. La propagación de grietas provocadas por fatiga en varias partes de un avión ha sido el tema de extensos estudios en años re-

cientes. Los datos adjuntos se componen de vidas de propagación (horas de vuelo/10^4) para alcanzar un tamaño de agrietamiento dado en orificios para sujetadores utilizados en aviones militares ("Statistical Crack Propagation in Fastener Holes Under Spectrum Loading", *J. Aircraft*, 1983: 1028–1032):

.736	.863	.865	.913	.915	.937	.983	1.007
1.011	1.064	1.109	1.132	1.140	1.153	1.253	1.394

a. Calcule y compare los valores de la media y mediana muestrales.
b. ¿En cuánto se podría disminuir la observación muestral más grande sin afectar el valor de la mediana?

40. Calcule la mediana muestral, media recortada 25%, media recortada 10% y media muestral de los datos de duración dados en el ejercicio 27 y compare estas medidas.

41. Se eligió una muestra de $n = 10$ automóviles y cada uno se sometió a una prueba de choque a 5 mph. Denotando un carro sin daños visibles con S y un carro con daños con F, los resultados fueron los siguientes:

S S F S S S F F S S

a. ¿Cuál es el valor de la proporción muestral de éxitos x/n?
b. Reemplace cada S con 1 y cada F con 0. Acto seguido calcule $\bar{x}$ de esta muestra numéricamente codificada. ¿Cómo se compara $\bar{x}$ con x/n?
c. Suponga que se decide incluir 15 carros más en el experimento. ¿Cuántos de éstos tendrían que ser S para dar $x/n = .80$ para toda la muestra de 25 carros?

42. **a.** Si se agrega una constante c a cada x_i en una muestra y se obtiene $y_i = x_i + c$, ¿cómo se relacionan la media y mediana muestrales de las y_i con la media y mediana muestrales de las x_i? Verifique sus conjeturas.
b. Si cada x_i se multiplica por una constante c y se obtiene $y_i = cx_i$, responda la pregunta del inciso (a). De nuevo, verifique sus conjeturas.

43. Un experimento para estudiar la duración (en horas) de un cierto tipo de componente implicaba poner diez componentes en operación y observarlos durante 100 horas. Ocho de ellos fallaron durante dicho periodo y se registraron las duraciones. Denote las duraciones de los dos componentes que continuaron funcionando después de 100 horas por 100+. Las observaciones muestrales resultantes fueron:

48 79 100+ 35 92 86 57 100+ 17 29

¿Cuáles de las medidas del centro discutidas en esta sección pueden ser calculadas y cuáles son los valores de dichas medidas? [*Nota*: se dice que los datos obtenidos con este experimento están "censurados a la derecha".]

1.4 Medidas de variabilidad

El reporte de una medida de centro da sólo información parcial sobre un conjunto o distribución de datos. Diferentes muestras o poblaciones pueden tener medidas idénticas de centro y aún diferir una de otra en otras importantes maneras. La figura 1.19 muestra gráficas de puntos de tres muestras con las mismas media y mediana, aunque el grado de dispersión en torno al centro es diferente para las tres muestras. La primera tiene la cantidad más grande de variabilidad, la tercera tiene la cantidad más pequeña y la segunda es intermedia con respecto a las otras dos en este aspecto.

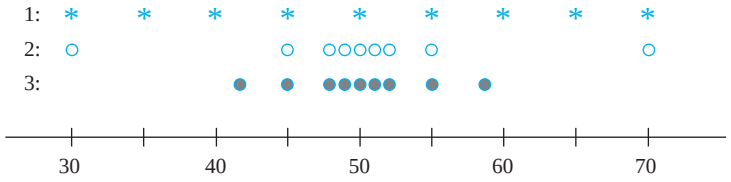


Figura 1.19 Muestras con medidas de centro idénticas pero diferentes cantidades de variabilidad

Medidas de variabilidad de datos muestrales

La medida más simple de variabilidad en una muestra es el **rango,** el cual es la diferencia entre los valores muestrales más grande y más pequeño. El valor del rango de la muestra 1 en la figura 1.19 es mucho más grande que el de la muestra 3, lo que refleja más variabilidad en la primera muestra que en la tercera. Un defecto del rango, no obstante, es que depende de sólo las dos observaciones más extremas y hace caso omiso de las posiciones de los $n - 2$ valores restantes. Las muestras 1 y 2 en la figura 1.19 tienen rangos idénti-

cos, aunque cuando se toman en cuenta las observaciones entre los dos extremos, existe mucho menos variabilidad o dispersión en la segunda muestra que en la primera.

Las medidas principales de variabilidad implican las **desviaciones de la media,** $x_1 - \bar{x}, x_2 - \bar{x}, \ldots, x_n - \bar{x}$. Es decir, las desviaciones de la media se obtienen restando $\bar{x}$ de cada una de la n observaciones muestrales. Una desviación será positiva si la observación es más grande que la media (a la derecha de la media sobre el eje de medición) y negativa si la observación es más pequeña que la media. Si todas las desviaciones son pequeñas en magnitud, entonces todas las x_i se aproximan a la media y hay poca variabilidad. Alternativamente, si algunas de las desviaciones son grandes en magnitud, entonces algunas x_i quedan lejos de $\bar{x}$ lo que sugiere una mayor cantidad de variabilidad. Una forma simple de combinar las desviaciones en una sola cantidad es promediarlas. Desafortunadamente, esto es una mala idea:

$$\text{suma de desviaciones} = \sum_{i=1}^{n}(x_i - \bar{x}) = 0$$

por lo que la desviación promedio siempre es cero. La verificación utiliza varias reglas estándar de la suma y el hecho de que $\sum \bar{x} = \bar{x} + \bar{x} + \cdots + \bar{x} = n\bar{x}$:

$$\sum(x_i - \bar{x}) = \sum x_i - \sum \bar{x} = \sum x_i - n\bar{x} = \sum x_i - n\left(\frac{1}{n}\sum x_i\right) = 0$$

¿Cómo se puede evitar que las desviaciones negativas y positivas se neutralicen entre sí cuando se combinan? Una posibilidad es trabajar con los valores absolutos de las desviaciones y calcular la desviación absoluta promedio $\sum |x_i - \bar{x}|/n$. Como la operación de valor absoluto conduce a un número de dificultades teóricas, considérense en cambio las desviaciones al cuadrado $(x_1 - \bar{x})^2, (x_2 - \bar{x})^2, \ldots, (x_n - \bar{x})^2$. En vez de utilizar la desviación al cuadrado promedio $\sum(x_i - \bar{x})^2/n$, por varias razones se divide la suma de desviaciones al cuadrado entre $n - 1$ en lugar de entre n.

DEFINICIÓN

La **varianza muestral,** denotada por s^2 está dada por

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n - 1} = \frac{S_{xx}}{n - 1}$$

La **desviación estándar muestral,** denotada por s, es la raíz cuadrada (positiva) de la varianza:

$$s = \sqrt{s^2}$$

Obsérvese que s^2 y s son no negativas. La unidad de s es la misma que la de cada una de las x_i. Si por ejemplo, las observaciones son eficiencias de combustible en millas por galón, entonces se podría tener $s = 2.0$ mpg. Una interpretación preliminar de la desviación estándar muestral es que es el tamaño de una desviación típica o representativa de la media muestral dentro de la muestra dada. Por tanto si $s = 2.0$ mpg, entonces algunas x_i en la muestra se aproximan más que 2.0 a $\bar{x}$, en tanto que otras están más alejadas; 2.0 es una desviación representativa (o "estándar") de la eficiencia de combustible media. Si $s = 3.0$ para una segunda muestra de carros de otro tipo, una desviación típica en esta muestra es aproximadamente 1.5 veces la de la primera muestra, una indicación de más variabilidad en la segunda muestra.

Ejemplo 1.17 El sitio web www.fueleconomy.gov contiene una gran cantidad de información acerca de las características del combustible de varios vehículos. Además de las calificaciones de millaje de la EPA, hay muchos vehículos para los que los usuarios han informado de sus propios valores de eficiencia de combustible (mpg). Considere la siguiente muestra de $n = 11$ eficiencias para el Ford Focus 2009 equipado con transmisión automática (para

este modelo, la EPA informa de una calificación general de 27 mpg-24 mpg en ciudad y 33 mpg en carretera):

Automóvil	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	
1	27.3	−5.96	35.522	
2	27.9	−5.36	28.730	
3	32.9	−0.36	0.130	
4	35.2	1.94	3.764	
5	44.9	11.64	135.490	
6	39.9	6.64	44.090	
7	30.0	−3.26	10.628	
8	29.7	−3.56	12.674	
9	28.5	−4.76	22.658	
10	32.0	−1.26	1.588	
11	37.6	4.34	18.836	
	$\sum x_i = 365.9$	$\sum (x_i - \bar{x}) = .04$	$\sum (x_i - \bar{x})^2 = 314.106$	$\bar{x} = 33.26$

Los efectos de redondeo hacen que la suma de las desviaciones no sea exactamente cero. El numerador de s^2 es $S_{xx} = 314.106$, por consiguiente

$$s^2 = \frac{S_{xx}}{n-1} = \frac{314.106}{11-1} = 31.41, \qquad s = 5.60$$

El tamaño de una desviación representativa de la media de la muestra 33.26 es de aproximadamente 5.6 mpg. Nota: de las nueve personas que también reportaron hábitos de conducción, sólo tres hicieron más del 80% de ésta en carretera; apostamos a que puede adivinar los coches que conducían. Todavía no tenemos idea de por qué los 11 valores registrados exceden la cifra de la EPA, tal vez sólo los conductores con una eficiencia de combustible realmente buena comuniquen sus resultados. ■

Motivación para s^2

Para explicar el porqué del divisor $n - 1$ en s^2, obsérvese primero que en tanto que s^2 mide la variabilidad muestral, existe una medida de variabilidad en la población llamada *varianza poblacional*. Se utilizará σ^2 (el cuadrado de la letra griega sigma minúscula) para denotar la varianza poblacional y σ para denotar la desviación estándar poblacional (la raíz cuadrada de σ^2). Cuando la población es finita y se compone de N valores,

$$\sigma^2 = \sum_{i=1}^{N} (x_i - \mu)^2 / N$$

la cual es el promedio de todas las desviaciones al cuadrado con respecto a la media poblacional (para la población, el divisor es N y no $N - 1$). En los capítulos 3 y 4 aparecen definiciones más generales de σ^2.

Así como $\bar{x}$ se utilizará para hacer inferencias sobre la media poblacional μ, se deberá definir la varianza muestral de modo que pueda ser utilizada para hacer inferencias sobre σ^2. Ahora obsérvese que σ^2 implica desviaciones cuadradas con respecto a la media poblacional μ. Si en realidad se conociera el valor de μ, entonces se podría definir la varianza muestral como la desviación al cuadrado promedio de las x_i de la muestra con respecto a μ. Sin embargo, el valor de μ casi nunca es conocido, por lo que se debe utilizar el cuadrado de la suma de las desviaciones con respecto a $\bar{x}$. Pero *las x_i tienden a acercarse más a su valor promedio $\bar{x}$ que el promedio poblacional μ, así que para compensar esto se utiliza el divisor $n - 1$ en lugar de n*. En otras palabras, si se utiliza un divisor n en la varianza muestral, entonces la cantidad resultante tendería a subestimar σ^2 (se producen valores demasiado pequeños en promedio), mientras que si se divide entre el divisor un poco más pequeño $n - 1$ se corrige esta subestimación.

Se acostumbra referirse a s^2 que está basada en $n - 1$ **grados de libertad** (gl). Esta terminología se deriva del hecho de que aunque s^2 está basada en las n cantidades $x_1 - \bar{x}, x_2 - \bar{x}, \ldots, x_n - \bar{x}$, éstas suman 0, por lo que al especificar los valores de cualquier $n - 1$ de las cantidades se determina el valor restante. Por ejemplo, si $n = 4$ y $x_1 - \bar{x} = 8, x_2 - \bar{x} = -6$ y $x_4 - \bar{x} = -4$ entonces automáticamente $x_3 - \bar{x} = 2$, así que sólo tres de los cuatro valores de $x_i - \bar{x}$ son libremente determinados (3 gl).

Una fórmula para calcular s^2

Es mejor obtener s^2 con software estadístico o bien utilizar una calculadora que permita ingresar datos en la memoria y luego ver s^2 con un solo golpe de tecla. Si su calculadora no tiene esta capacidad, existe una fórmula alternativa para S_{xx} que evita calcular las desviaciones. La fórmula implica $(\sum x_i)^2$, sumar y luego elevar al cuadrado, y $\sum x_i^2$, elevar al cuadrado y sumar.

Una *expresión* alternativa para el numerador de s^2 es

$$S_{xx} = \sum(x_i - \bar{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

Demostración Como $\bar{x} = \sum x_i/n$, $n\bar{x}^2 = (\sum x_i)^2/n$. Entonces,

$$\begin{aligned}\sum(x_i - \bar{x})^2 &= \sum(x_i^2 - 2\bar{x} \cdot x_i + \bar{x}^2) = \sum x_i^2 - 2\bar{x}\sum x_i + \sum(\bar{x})^2 \\ &= \sum x_i^2 - 2\bar{x} \cdot n\bar{x} + n(\bar{x})^2 = \sum x_i^2 - n(\bar{x})^2\end{aligned}$$

Ejemplo 1.18 La luxación traumática de rodilla a menudo requiere cirugía para reparar los ligamentos rotos. Una medida de la recuperación es la amplitud de movimiento (medido como el ángulo formado cuando, a partir de la pierna estirada, la rodilla se dobla en la medida de lo posible). Los datos que figuran en el rango de movimiento posquirúrgico aparecieron en el artículo "Reconstruction of the Anterior and Posterior Cruciate Ligaments After Knee Dislocation" (*Amer. J. Sports Med.*, 1999: 189–197):

154 142 137 133 122 126 135 135 108 120 127 134 122

La suma de estas 13 muestras observadas es $\sum x_i = 1695$, y la suma de sus cuadrados es

$$\sum x_i^2 = (154)^2 + (142)^2 + \cdots + (122)^2 = 222{,}581$$

Por tanto, el numerador de la varianza muestral es

$$S_{xx} = \sum x_i^2 - [(\sum x_i)^2]/n = 222{,}581 - (1695)^2/13 = 1579.0769$$

de donde $s^2 = 1579.0769/12 = 131.59$ y $s = 11.47$. ■

Tanto la fórmula de la definición y la fórmula de cálculo para s^2 pueden ser sensibles al redondeo, por lo que en los cálculos intermedios se debe utilizar la mayor precisión decimal posible.

Varias propiedades de s^2 pueden mejorar la comprensión y facilitar el cálculo.

PROPOSICIÓN

Sean $x_1, x_2, \ldots, x_n$ una muestra y c cualquier constante diferente de cero.

1. Si $y_1 = x_1 + c, y_2 = x_2 + c, \ldots, y_n = x_n + c$, entonces $s_y^2 = s_x^2$ y

2. Si $y_1 = cx_1, \ldots, y_n = cx_n$, entonces $s_y^2 = c^2 s_x^2$, $s_y = |c|s_x$

donde s_x^2 es la varianza muestral de las x y s_y^2 es la varianza muestral de las y.

En palabras, el resultado 1 dice que si se suma (o resta) una constante c de cada valor de dato, la varianza no cambia. Esto es intuitivo, puesto que la adición o sustracción de c cambia la ubicación del conjunto de datos pero deja inalteradas las distancias entre los valores de datos. De acuerdo con el resultado 2, la multiplicación de cada x_i por c hace que s^2 sea multiplicada por un factor de c^2. Estas propiedades pueden ser comprobadas al observar en el resultado 1 que $\bar{y} = \bar{x} + c$ y en el resultado 2 que $\bar{y} = c\bar{x}$.

Gráficas de caja

Las gráficas de tallo y hojas e histogramas transmiten impresiones un tanto generales sobre un conjunto de datos, mientras que un resumen único tal como la media o la desviación estándar se enfoca en sólo un aspecto de los datos. En años recientes se ha utilizado con éxito un resumen gráfico llamado *gráfica de caja* para describir varias de las características más prominentes de un conjunto de datos. Estas características incluyen (1) el centro, (2) la dispersión, (3) el grado y naturaleza de cualquier alejamiento de la simetría y (4) la identificación de las observaciones "extremas o apartadas" inusualmente alejadas del cuerpo principal de los datos. Como incluso un solo valor extremo puede afectar drásticamente los valores de $\bar{x}$ y s, una gráfica de caja está basada en medidas "resistentes" a la presencia de unos cuantos valores apartados: la mediana y una medida de variabilidad llamada *dispersión de los cuartos*.

DEFINICIÓN

Se ordenan las n observaciones de la más pequeña a la más grande y se separa la mitad más pequeña de la más grande; se incluye la mediana $\tilde{x}$ en ambas mitades si n es impar. En tal caso el **cuarto inferior** es la mediana de la mitad más pequeña y el **cuarto superior** es la mediana de la mitad más grande. Una medida de dispersión que es resistente a los valores apartados es la **dispersión de los cuartos** f_s, dada por

$$f_s = \text{cuarto superior} - \text{cuarto inferior}$$

En general, la dispersión de los cuartos no se ve afectada por las posiciones de las observaciones comprendidas en el 25% más pequeño o el 25% más grande de los datos. Por consiguiente es resistente a valores apartados.

La gráfica de caja más simple se basa en el siguiente resumen de cinco números:

x_i más pequeñas cuarto inferior mediana cuarto superior x_i más grandes

Primero, se traza una escala de medición horizontal. Luego se coloca un rectángulo sobre este eje; el lado izquierdo del rectángulo está en el cuarto inferior y el derecho en el cuarto superior (por lo que el ancho de la caja $= f_s$). Se coloca un segmento de línea vertical o algún otro símbolo adentro del rectángulo en la ubicación de la mediana; la posición del símbolo de mediana con respecto a los dos lados da información sobre asimetría en el 50% medio de los datos. Por último, se trazan "bigotes" hacia fuera de ambos extremos del rectángulo hacia las observaciones más pequeñas y más grandes. También se puede trazar una gráfica de caja con orientación vertical mediante modificaciones obvias en el proceso de construcción.

Ejemplo 1.19 Se utilizó ultrasonido para reunir los datos adjuntos de corrosión en el espesor de la placa de piso de un tanque elevado utilizado para almacenar petróleo crudo ("Statistical Analysis of UT Corrosion Data from Floor Plates of a Crude Oil Aboveground Storage Tank", *Materials Eval*., 1994: 846–849); cada observación es la profundidad de la picadura más grande en la placa, expresada en milésimas de pulgada

40 52 55 60 70 75 85 85 90 90 92 94 94 95 98 100 115 125 125

El resumen de cinco números es como sigue:

x_i más pequeña = 40 cuarto inferior = 72.5 $\tilde{x}$ = 90 cuarto superior = 96.5
x_i más grande = 125

La figura 1.20 muestra la gráfica de caja resultante. El lado derecho de la caja está mucho más cerca a la mediana que el izquierdo, lo que indica una simetría sustancial en la mitad central de los datos. El ancho de la caja (f_s) también es razonablemente grande con respecto al rango de datos (distancia entre las puntas de los bigotes).

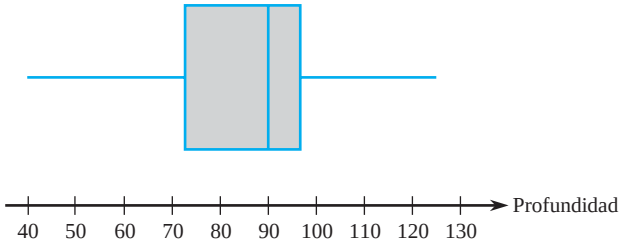


Figura 1.20 Gráfica de caja para los datos de corrosión

La figura 1.20 muestra los resultados obtenidos con Minitab en respuesta a la petición de describir los datos de corrosión. Q1 y Q3 son los cuartiles inferior y superior; éstos son similares a los cuartos pero se calculan de una manera un poco diferente; la media SE es $s/\sqrt{n}$; ésta será una importante cantidad en el trabajo subsiguiente con respecto a inferencias en torno a μ.

```
Variable        N      Mean   Median   TrMean   StDev   SE Mean
   depth       19     86.32    90.00    86.76   23.32      5.35

Variable   Minimum   Maximum       Q1       Q3
   depth     40.00    125.00    70.00    98.00
```

Figura 1.21 Descripción Minitab de los datos de la profundidad del pozo ■

Gráficas de caja que muestran valores apartados

Una gráfica de caja puede ser embellecida para indicar explícitamente la presencia de valores apartados. Muchos procedimientos inferenciales se basan en la suposición de que la distribución de la población es normal (un cierto tipo de curva en forma de campana). Incluso un solo valor apartado extremo que aparezca en la muestra advierte al investigador que tales procedimientos pueden ser no confiables y la presencia de varios valores apartados moderados transmite el mismo mensaje.

DEFINICIÓN

Cualquier observación a más de $1.5f_s$ del cuarto más cercano es un **valor apartado.** Un valor apartado es **extremo** si se encuentra a más de $3f_s$ del cuarto más cercano, y **moderado** en caso contrario.

Modifíquese ahora la construcción previa de una gráfica de caja trazando un bigote que sale de cada extremo de la caja hacia las observaciones más pequeñas y más grandes que *no* son valores apartados. Cada valor apartado moderado está representado por un círculo cerrado y cada valor apartado extremo por uno abierto. Algunos programas de computadora estadísticos no distinguen entre valores apartados moderados y extremos.

Ejemplo 1.20 La ley Clean Water (agua limpia) y las modificaciones posteriores requieren que todas las aguas en Estados Unidos alcancen los objetivos de reducción de la contaminación para garantizar que el agua sea “apta para la pesca y para nadar”. El artículo “Spurious Correlation in the USEPA Rating Curve Method for Estimating Pollutant Loads” (*J. of*

Environ. Engr., 2008: 610–618) ha investigado diferentes técnicas para estimar las cargas contaminantes en las cuencas hidrográficas; los autores "discuten la necesidad imperiosa del uso racional de los métodos estadísticos" para este fin. Entre los datos que se consideran está la siguiente muestra de cargas de NT (nitrógeno total) (kg N/día) a partir de una determinada ubicación en la Bahía de Chesapeake, que aparecen aquí en orden creciente.

9.69	13.16	17.09	18.12	23.70	24.07	24.29	26.43
30.75	31.54	35.07	36.99	40.32	42.51	45.64	48.22
49.98	50.06	55.02	57.00	58.41	61.31	64.25	65.24
66.14	67.68	81.40	90.80	92.17	92.42	100.82	101.94
103.61	106.28	106.80	108.69	114.61	120.86	124.54	143.27
143.75	149.64	167.79	182.50	192.55	193.53	271.57	292.61
312.45	352.09	371.47	444.68	460.86	563.92	690.11	826.54
1529.35							

El resumen de las cantidades pertinentes es

$$\tilde{x} = 92.17 \qquad 4^{o} \text{ inferior} = 45.64 \qquad 4^{o} \text{ superior} = 167.79$$

$$f_s = 122.15 \qquad 1.5f_s = 183.225 \qquad 3f_s = 366.45$$

Restando $1.5f_s$ del 4º inferior da un número negativo y ninguna de las observaciones es negativa, así que no hay valores atípicos en el extremo inferior de los datos. Sin embargo,

$$4^{o} \text{ superior} + 1.5f_s = 351.015 \quad 4^{o} \text{ superior} + 3f_s = 534.24$$

Por tanto las cuatro observaciones más grandes, 563.92, 690.11, 826.54 y 1529.35, son valores apartados extremos; y 352.09, 371.47, 444.68, y 460.86 son valores apartados moderados.

Los bigotes en la gráfica de caja de la figura 1.22 se extienden hacia afuera de la observación más pequeña, 9.69, en el extremo inferior, y 312.45, la observación más grande en el extremo superior que no es un valor apartado. Hay una cierta asimetría positiva en la mitad central de los datos (la línea mediana está un poco más cerca del borde izquierdo de la caja que del extremo derecho) y, en general una gran asimetría positiva.

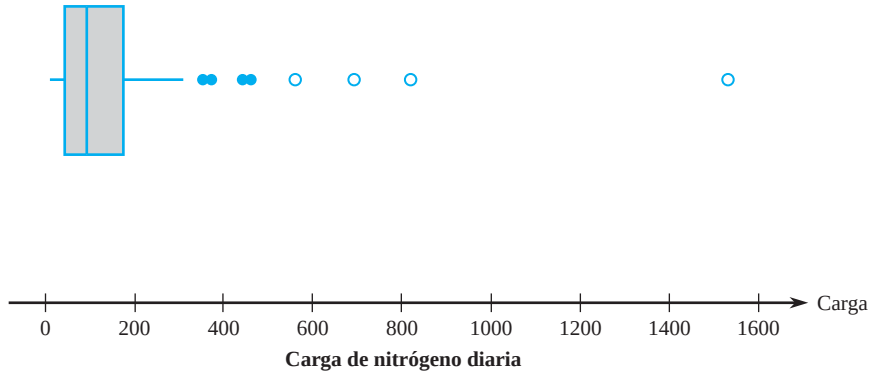


Figura 1.22 Gráfica de caja de los datos de la carga de nitrógeno mostrando los valores apartados moderados y extremos

Gráficas de caja comparativas

Una gráfica de caja comparativa o lado a lado es una forma muy efectiva de revelar similitudes y diferencias entre dos o más conjuntos de datos compuestos de observaciones de la misma variable, observaciones de eficiencia de consumo de combustible de cuatro tipos distintos de automóviles, rendimientos de cosechas de tres variedades diferentes, y así sucesivamente.

Ejemplo 1.21 En años recientes, algunas evidencias sugieren que las altas concentraciones de radón bajo techo pueden estar ligadas al desarrollo de cánceres en niños, pero muchos profesionales de la salud aún no están convencidos. Un artículo reciente ("Indoor Radon and Childhood Cancer", *The Lancet*, 1991: 1537–1538) presentó los datos adjuntos sobre concentración de radón (Bq/m^3) en dos muestras diferentes de casas. La primera consistió en casas en las cuales un niño diagnosticado con cáncer había estado residiendo. Las casas en la segunda muestra no incluían casos registrados de cáncer infantil. La figura 1.23 presenta una gráfica de tallo y hojas de los datos.

1. Con cáncer		2. Sin cáncer
9683795	0	95768397678993
86071815066815233150	1	12271713114
12302731	2	99494191
8349	3	839
5	4	
7	5	55
	6	
	7	Tallo: dígito de las decenas
HI: 210	8	5 Hoja: dígito de las unidades

Figura 1.23 Gráfica de tallos y hojas para el ejemplo 1.21

El resumen de cantidades numéricas es el siguiente:

	$\bar{x}$	$\tilde{x}$	s	f_s
Con cáncer	22.8	16.0	31.7	11.0
Sin cáncer	19.2	12.0	17.0	18.0

Los valores tanto de la media como de la mediana sugieren que la muestra con cáncer se encuentra en el centro un poco a la derecha de la muestra sin cáncer sobre la escala de medición. La media, sin embargo, exagera la magnitud de este desplazamiento, en gran medida debido a la observación 210 en la muestra con cáncer. Los valores de s sugieren más variabilidad en la muestra con cáncer que en la muestra sin cáncer, pero las dispersiones de los cuartos contradicen esta impresión. De nuevo, la observación 210, un valor apartado extremo, es el culpable. La figura 1.24 muestra una gráfica de caja comparativa generada por el programa de computadora S-Plus. La caja sin cáncer aparece alargada en compara-

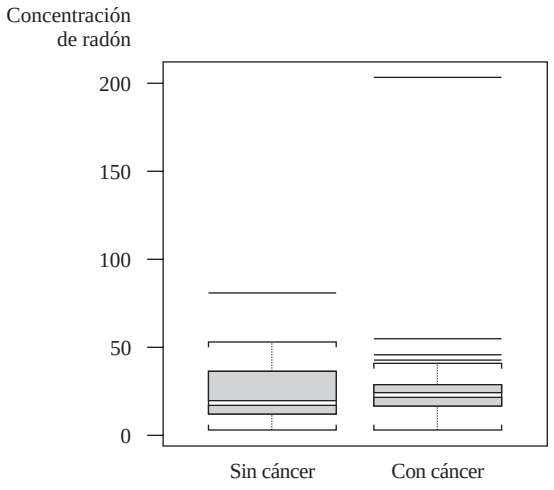


Figura 1.24 Gráfica de caja para el ejemplo 1.21, tomada de S-plus

ción con la caja con cáncer ($f_s = 18$ vs. $f_s = 11$) y las posiciones de las líneas medianas en las dos cajas muestran más asimetría en la mitad media de la muestra sin cáncer que la muestra con cáncer. Los valores apartados están representados por segmentos de línea horizontales y no hay distinción entre los valores apartados moderados y extremos. ■

EJERCICIOS Sección 1.4 (44–61)

44. El artículo "Oxygen Consumption During Fire Suppression: Error of Heart Rate Estimation" (*Ergonomics*, 1991: 1469–1474) reportó los siguientes datos sobre consumo de oxígeno (mL/kg/min) para una muestra de diez bomberos que realizaron un simulacro de supresión de incendio.

29.5 49.3 30.6 28.2 28.0 26.3 33.9 29.4 23.5 31.6

Calcule lo siguiente

a. El rango muestral
b. La varianza muestral s^2 a partir de la definición (es decir, calculando primero las desviaciones y luego elevándolas al cuadrado, etc.)
c. La desviación estándar muestral
d. s^2 utilizando el método más corto

45. Se determinó el valor del módulo de Young (GPa) de placas fundidas compuestas de ciertos sustratos intermetálicos y se obtuvieron las siguientes observaciones muestrales ("Strength and Modulus of a Molybdenum-Coated Ti-25A1-10Nb-3U-1Mo Intermetallic", *J. of Materials Engr. and Performance*, 1997: 46-50:

116.4 115.9 114.6 115.2 115.8

a. Calcule $\bar{x}$ y las desviaciones de la media.
b. Use las desviaciones calculadas en el inciso (a) para obtener la varianza muestral y la desviación estándar muestral.
c. Calcule s^2 utilizando la fórmula computacional para el numerador S_{xx}.
d. Reste 100 de cada observación para obtener una muestra de valores transformados. Ahora calcule la varianza muestral de estos valores transformados y compárela con s^2 de los datos originales.

46. Las observaciones adjuntas de viscosidad estabilizada (cP) realizadas en muestras de un cierto grado de asfalto con 18% de caucho agregado se tomaron del artículo "Viscosity Characteristics of Rubber-Modified Asphalts" (*J. of Materials in Civil Engr.*, 1996: 153–156):

2781 2900 3013 2856 2888

a. ¿Cuáles son los valores de la media y mediana muestrales?
b. Calcule la varianza muestral por medio de la fórmula de cálculo. [*Sugerencia*: primero reste un número conveniente de cada observación.]

47. Calcule e interprete los valores de la mediana muestral, la media muestral y la desviación estándar muestral de las siguientes observaciones de resistencia a la fractura (MPa, leídas en una gráfica que aparece en el artículo "Heat-Resistant Active Brazing of Silicon Nitride: Mechanical Evaluation of Braze Joints", *Welding J.*, agosto de 1997):

87 93 96 98 105 114 128 131 142 168

48. El ejercicio 34 presentó los siguientes datos sobre concentración de endotoxina en polvo asentado obtenidos con una muestra de casas urbanas y una muestra de casas campestres:

U: 6.0 5.0 11.0 33.0 4.0 5.0 80.0 18.0 35.0 17.0 23.0
C: 4.0 14.0 11.0 9.0 9.0 8.0 4.0 20.0 5.0 8.9 21.0
9.2 3.0 2.0 0.3

a. Determine el valor de la desviación estándar muestral de cada muestra, interprete estos valores y luego contraste la variabilidad en las dos muestras. [*Sugerencia*: $\sum x_i = 237.0$ para la muestra urbana y 128.4 para la muestra campestre y $\sum x_i^2 = 10{,}079$ para la muestra urbana y 1617.94 para la muestra campestre.]
b. Calcule la dispersión de los cuartos de cada muestra y compare. ¿Las dispersiones de los cuartos transmiten el mismo mensaje sobre la variabilidad que las desviaciones estándar? Explique.
c. Los autores del artículo citado también proporcionan concentraciones de endotoxina en el polvo presente en bolsas captadoras de polvo:

U: 34.0 49.0 13.0 33.0 24.0 24.0 35.0 104.0 34.0 40.0 38.0 1.0
C: 2.0 64.0 6.0 17.0 35.0 11.0 17.0 13.0 5.0 27.0 23.0
28.0 10.0 13.0 0.2

Construya una gráfica de caja comparativa (como se hizo en el artículo citado) y compare y contraste las cuatro muestras.

49. Un estudio de la relación entre edad y varias funciones visuales (tales como agudeza y percepción de profundidad) reportó las siguientes observaciones en el área de la lámina esclerótica (mm^2) de las cabezas del nervio óptico humano ("Morphometry of Nerve Fiber Bundle Pores in the Optic Nerve Head of the Human", *Experimental Eye Research*, 1988: 559–568):

2.75 2.62 2.74 3.85 2.34 2.74 3.93 4.21 3.88
4.33 3.46 4.52 2.43 3.65 2.78 3.56 3.01

a. Calcule $\sum x_i$ y $\sum x_i^2$.
b. Use los valores calculados en el inciso (a) para calcular la varianza muestral s^2 y luego la desviación estándar muestral s.

50. En 1997 una mujer demandó a un fabricante de teclados de computadora y lo acusó de que sus repetidas lesiones por esfuerzo eran provocadas por el teclado (*Genessy v. Digital Equipment Corp.*). El jurado le adjudicó $3.5 millones por el

dolor y sufrimiento pero la corte anuló dicha adjudicación por considerarla una compensación irrazonable. Al hacer esta determinación, la corte identificó un grupo "normativo" de 27 casos similares y especificó que una adjudicación razonable estaría dentro de dos desviaciones estándar de la media de las adjudicaciones en los 27 casos. Las 27 adjudicaciones fueron (en el rango de los $1000) 37, 60, 75, 115, 135, 140, 149, 150, 238, 290, 340, 410, 600, 750, 750, 750, 1050, 1100, 1139, 1150, 1200, 1200, 1250, 1576, 1700, 1825 y 2000, con las cuales $\sum x_i = 20{,}179$, $\sum x_i^2 = 24{,}657{,}511$. ¿Cuál es la cantidad máxima posible que podría ser adjudicada conforme a la regla de dos desviaciones estándar?

51. El artículo "A Thin-Film Oxygen Uptake Test for the Evaluation of Automotive Crankcase Lubricants" (*Lubric. Engr.*, 1984: 75–83) reportó los siguientes datos sobre tiempo de inducción de oxidación (min) de varios aceites comerciales:

87 103 130 160 180 195 132 145 211 105 145
153 152 138 87 99 93 119 129

a. Calcule la varianza y la desviación estándar muestrales.
b. Si las observaciones se volvieran a expresar en horas, ¿cuáles serían los valores resultantes de la varianza de la muestra y la desviación estándar muestral? Responda sin realizar en realidad la reexpresión.

52. Las primeras cuatro desviaciones de la media en una muestra de $n = 5$ tiempos de reacción fueron .3, .9, 1.0 y 1.3. ¿Cuál es la quinta desviación de la media? Dé una muestra para la cual éstas son las cinco desviaciones de la media.

53. Un **fondo mutuo** es un esquema de inversiones administrado por profesionales que invierten el dinero de muchos inversionistas en una variedad de valores. Los fondos de crecimiento se centran principalmente en el aumento del valor de las inversiones, mientras que los fondos mezclados buscan un equilibrio entre ingresos corrientes y el crecimiento. Aquí hay datos sobre la proporción de gastos (gastos en % de los activos, de www.morningstar.com) para las muestras de los 20 fondos de gran capitalización equilibrada y 20 fondos de crecimiento de gran capitalización ("de gran capitalización" se refiere a los tamaños de las empresas en las que los fondos se invierten; los tamaños de la población son 825 y 762, respectivamente):

Mez	1.03	1.23	1.10	1.64	1.30
	1.27	1.25	0.78	1.05	0.64
	0.94	2.86	1.05	0.75	0.09
	0.79	1.61	1.26	0.93	0.84
Cr	0.52	1.06	1.26	2.17	1.55
	0.99	1.10	1.07	1.81	2.05
	0.91	0.79	1.39	0.62	1.52
	1.02	1.10	1.78	1.01	1.15

a. Calcule y compare los valores de $\bar{x}$, $\tilde{x}$ y s para los dos tipos de fondos.
b. Construya una gráfica de caja comparativa para los dos tipos de fondos y comente acerca de las características interesantes.

54. El agarre se aplica para producir fuerzas superficiales normales que comprimen el objeto que se quiere aferrar. Los ejemplos incluyen a dos personas dándose la mano, o una enfermera apretando el antebrazo del paciente para detener el sangrado. El artículo "Investigation of Grip Force, Normal Force, Contact Area, Hand Size, and Handle Size for Cylindrical Handles" (*Human Factors*, 2008: 734–744) incluye los siguientes datos sobre la fuerza de prensión (N) para una muestra de 42 individuos:

16 18 18 26 33 41 54 56 66 68 87 91 95
98 106 109 111 118 127 127 135 145 147 149 151 168
172 183 189 190 200 210 220 229 230 233 238 244 259
294 329 403

a. Construya un diagrama de tallo y hojas sobre la base de repetir cada valor de tallo dos veces y comente sobre las características interesantes.
b. Determine los valores de los cuartos y el cuarto disperso.
c. Construya una gráfica de caja basada en el resumen de cinco números y comente sobre sus características.
d. ¿Qué tan grande o pequeña tiene que ser una observación para calificar como valor apartado? ¿Como valor apartado extremo? ¿Hay valores apartados?
e. ¿Por cuánto podría disminuir la observación 403, actualmente la más grande, sin afectar f_s?

55. He aquí una gráfica de tallo y hojas de los datos de tiempo de escape introducidos en el ejercicio 36 de este capítulo.

32	55
33	49
34	
35	6699
36	34469
37	03345
38	9
39	2347
40	23
41	
42	4

a. Determine el valor de la dispersión de los cuartos.
b. ¿Hay algunos valores apartados en la muestra? ¿Algunos valores apartados extremos?
c. Construya una gráfica de caja y comente sobre sus características.
d. ¿En cuánto se podría disminuir la observación más grande, actualmente de 424, sin afectar el valor de la dispersión de los cuartos?

56. Los siguientes datos sobre el contenido de alcohol destilado (%) para una muestra de 35 vinos de Oporto fue extraído del artículo "A Method for the Estimation of Alcohol in Fortified Wines Using Hydrometer Baumé and Refractometer Brix" (*Amer. J. Enol. Vitic.*, 2006: 486–490). Cada valor es un promedio de dos medidas por duplicado.

16.35 18.85 16.20 17.75 19.58 17.73 22.75 23.78 23.25
19.08 19.62 19.20 20.05 17.85 19.17 19.48 20.00 19.97
17.48 17.15 19.07 19.90 18.68 18.82 19.03 19.45 19.37
19.20 18.00 19.60 19.33 21.22 19.50 15.30 22.25

Utilice los métodos de este capítulo, incluyendo un diagrama de caja que muestre los valores atípicos, para describir y resumir los datos.

57. Se seleccionó una muestra de 20 botellas de vidrio de un tipo particular y se determinó la resistencia de cada botella a la presión interna. Considere la siguiente información parcial sobre la muestra:

mediana = 202.2 cuarto inferior = 196.0
cuarto superior = 216.8

Las tres observaciones más pequeñas	125.8	188.1	193.7
Las tres observaciones más grandes	221.3	230.5	250.2

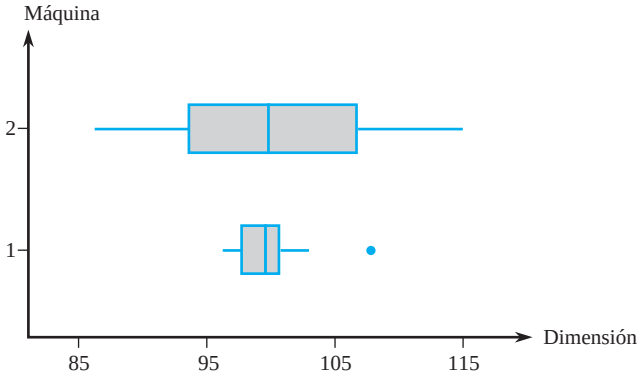
a. ¿Hay valores apartados en la muestra? ¿Algunos valores apartados extremos?
b. Construya una gráfica de caja que muestre los valores apartados y comente sobre cualesquiera características interesantes.

58. Una compañía utiliza dos máquinas diferentes para fabricar piezas de cierto tipo. Durante un solo turno, se obtuvo una muestra de $n = 20$ piezas producidas por cada máquina y se determinó el valor de una dimensión crítica particular de cada pieza. La gráfica de caja comparativa que aparece en la parte inferior de esta página se construyó con los datos resultantes. Compare y contraste las dos muestras.

59. Se determinó la concentración de cocaína (mg/L) con una muestra de individuos que murieron de delirio excitado (DE) inducido por el consumo de cocaína y con una muestra de aquellos que murieron de una sobredosis de cocaína sin delirio excitado; el tiempo de sobrevivencia de las personas en ambos grupos fue a lo sumo de 6 horas. Los datos adjuntos se tomaron de una gráfica de caja comparativa incluida en el artículo "Fatal Excited Delirium Following Cocaine Use" (*J. of Forensic Sciences*, 1997: 25–31).

Con DE	0	0	0	0	.1	.1	.1	.1	.2	.2	.3	.3
	.3	.4	.5	.7	.8	1.0	1.5	2.7	2.8			
	3.5	4.0	8.9	9.2	11.7	21.0						
Sin DE	0	0	0	0	0	.1	.1	.1	.1	.2	.2	.2
	.3	.3	.3	.4	.5	.5	.6	.8	.9	1.0		
	1.2	1.4	1.5	1.7	2.0	3.2	3.5	4.1				
	4.3	4.8	5.0	5.6	5.9	6.0	6.4	7.9				
	8.3	8.7	9.1	9.6	9.9	11.0	11.5					
	12.2	12.7	14.0	16.6	17.8							

Gráfica de caja comparativa para el ejercicio 58



a. Determine las medianas, cuartos y dispersiones de los cuartos de las dos muestras.
b. ¿Existen algunos valores apartados en una u otra muestra? ¿Algunos valores apartados extremos?
c. Construya una gráfica de caja comparativa y utilícela como base para comparar y contrastar las muestras con DE y sin DE.

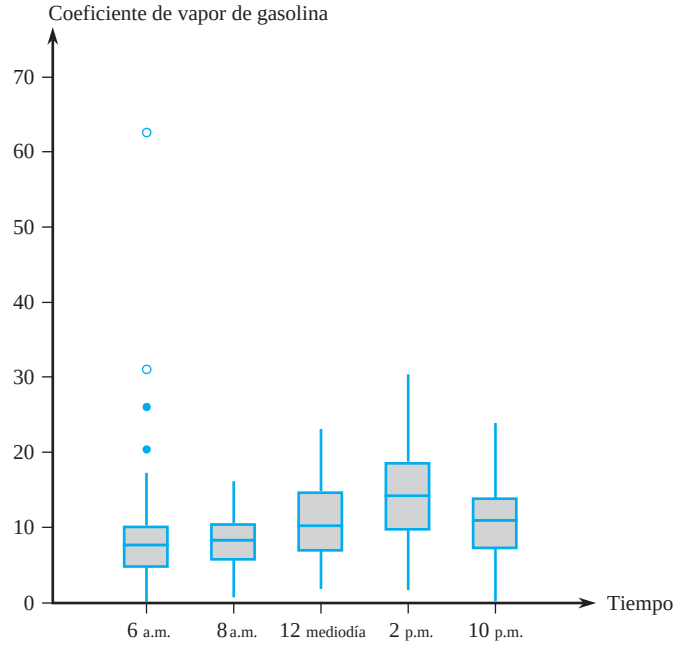
60. Se obtuvieron observaciones de resistencia al estallamiento ($lb/pulg^2$) con pruebas tanto con soldaduras de cierre de tobera como con soldaduras para tobera de envases de producción ("Proper Procedures Are the Key to Welding Radioactive Waste Cannisters", *Welding J.*, agosto de 1997: 61–67).

Prueba	7200	6100	7300	7300	8000	7400
	7300	7300	8000	6700	8300	
Envase	5250	5625	5900	5900	5700	6050
	5800	6000	5875	6100	5850	6600

Construya una gráfica de caja comparativa y comente sobre las características interesantes (el artículo citado no incluía tal gráfica, pero los autores comentaron que habían visto una.)

61. La gráfica de caja comparativa adjunta de coeficientes de vapor de gasolina para vehículos en Detroit apareció en el artículo "Receptor Modeling Approach to VOC Emission Inventory Validation" (*J. of Envir. Engr.*, 1995: 483–490). Discuta las características interesantes.

Gráfica de caja comparativa para el ejercicio 61



EJERCICIOS SUPLEMENTARIOS (62–83)

62. Considere la siguiente información sobre resistencia a la tensión final (lb/pulg) de una muestra de $n = 4$ probetas de alambre de cobre al zirconio duro (de "Characterization Methods for Fine Copper Wire", *Wire J. Intl.*, agosto de 1997: 74–80):

$\bar{x} = 76{,}831 \quad s = 180 \quad x_i$ más pequeña $= 76{,}683$
x_i más grande $= 77{,}048$

Determine los valores de las dos observaciones muestrales intermedias (¡pero no lo haga mediante conjeturas sucesivas!)

63. Se tomó una muestra de 77 personas que trabajan en una oficina particular y se determinó el nivel de ruido (dBA) experimentado por cada individuo, dando los siguientes datos ("Acceptable Noise Levels for Construction Site Offices", *Building Serv. Engr. Research and Technology*, 2009: 87–94).

55.3	55.3	55.3	55.9	55.9	55.9	55.9	56.1	56.1	56.1	56.1
56.1	56.1	56.8	56.8	57.0	57.0	57.0	57.8	57.8	57.8	57.9
57.9	57.9	58.8	58.8	58.8	59.8	59.8	59.8	62.2	62.2	63.8
63.8	63.8	63.9	63.9	63.9	64.7	64.7	64.7	65.1	65.1	65.1
65.3	65.3	65.3	65.3	67.4	67.4	67.4	67.4	68.7	68.7	68.7
68.7	69.0	70.4	70.4	71.2	71.2	71.2	73.0	73.0	73.1	73.1
74.6	74.6	74.6	74.6	79.3	79.3	79.3	79.3	83.0	83.0	83.0

Use algunos de los métodos estudiados en este capítulo para organizar, describir y resumir estos datos.

64. La corrosión por fricción es un proceso de desgaste que resulta de los movimientos oscilatorios tangenciales de pequeña amplitud en las piezas de una máquina. El artículo "Grease Effect on Fretting Wear of Mild Steel" (*Industrial Lubrication and Tribology*, 2008: 67–78) incluye los siguientes datos sobre el desgaste de volumen (10^{-4}mm^3) para los aceites base que tienen cuatro diferentes viscosidades.

Viscosidad	Desgaste					
20.4	58.8	30.8	27.3	29.9	17.7	76.5
30.2	44.5	47.1	48.7	41.6	32.8	18.3
89.4	73.3	57.1	66.0	93.8	133.2	81.1
252.6	30.6	24.2	16.6	38.9	28.7	23.6

a. El *coeficiente de variación muestral* $100s/\bar{x}$) evalúa el grado de variabilidad con respecto a la media (específicamente, la desviación estándar como porcentaje de la media). Calcule el coeficiente de variación para la muestra en cada viscosidad. Después, compare los resultados y coméntelos.
b. Construya una gráfica de caja comparativa de los datos y comente las características interesantes.

65. La distribución de frecuencia adjunta de observaciones de resistencia a la fractura (MPa) de barras de cerámica cocidas en un horno particular apareció en el artículo "Evaluating Tunnel Kiln Performance" (*Amer. Ceramic Soc. Bull.*, agosto de 1997: 59–63).

Clase	81−<83	83−<85	85−<87	87−<89	89−<91
Frecuencia	6	7	17	30	43
Clase	91−<93	93−<95	95−<97	97−<99	
Frecuencia	28	22	13	3	

a. Construya un histograma basado en frecuencias relativas y comente sobre cualesquiera características interesantes.
b. ¿Qué proporción de las observaciones de resistencia son por lo menos de 85? ¿Menores que 95?
c. Aproximadamente, ¿qué proporción de las observaciones son menores que 90?

66. Una deficiencia del microelemento selenio en la dieta puede impactar negativamente el crecimiento, la inmunidad, la función muscular y neuromuscular, y la fertilidad. La introducción de suplementos de selenio en vacas lecheras se justifica cuando las pasturas contienen niveles bajos del elemento. Los autores del artículo "Effects of Short-Term Supplementation with Selenised Yeast on Milk Production and Composition of Lactating Cows" (*Australian J. of Dairy Tech.*, 2004: 199–203) suministraron los siguientes datos sobre la concentración de selenio en la leche (mg/L) obtenidos con una muestra de vacas a las que se les administró un suplemento de selenio y una muestra de control de vacas a las que no se les administró suplemento, tanto inicialmente como después de un periodo de 9 días.

Obs	Se inicial	Contenido inicial	Se final	Contenido final
1	11.4	9.1	138.3	9.3
2	9.6	8.7	104.0	8.8
3	10.1	9.7	96.4	8.8
4	8.5	10.8	89.0	10.1
5	10.3	10.9	88.0	9.6
6	10.6	10.6	103.8	8.6
7	11.8	10.1	147.3	10.4
8	9.8	12.3	97.1	12.4
9	10.9	8.8	172.6	9.3
10	10.3	10.4	146.3	9.5
11	10.2	10.9	99.0	8.4
12	11.4	10.4	122.3	8.7
13	9.2	11.6	103.0	12.5
14	10.6	10.9	117.8	9.1
15	10.8		121.5	
16	8.2		93.0	

a. ¿Parecen ser similares las concentraciones iniciales de Se en las muestras de suplemento y en las de control? Use varias técnicas de este capítulo para resumir los datos y responder la pregunta planteada.
b. De nuevo use métodos de este capítulo para resumir los datos y luego describa cómo los valores de concentración de Se finales en el grupo de tratamiento difieren de aquellos en el grupo de control.

67. *Estenosis aórtica* se refiere al estrechamiento de la válvula aórtica en el corazón. El artículo "Correlation Analysis of Stenotic Aortic Valve Flow Patterns Using Phase Contrast MRI" (*Annals of Biomed. Engr.*, 2005: 878–887) dio los siguientes datos sobre el diámetro de la raíz aórtica (cm) y el género de una muestra de pacientes con varios grados de estenosis aórtica:

H: 3.7 3.4 3.7 4.0 3.9 3.8 3.4 3.6 3.1 4.0 3.4 3.8 3.5
M: 3.8 2.6 3.2 3.0 4.3 3.5 3.1 3.1 3.2 3.0

a. Compare y contraste los diámetros observados en los dos géneros.

b. Calcule una media recortada 10% de cada una de las dos muestras y compare las demás medidas del centro (de la muestra de hombre, se debe utilizar el método de interpolación mencionado en la sección 1.3).

68. a. ¿Con qué valor de c es mínima la cantidad $\sum(x_i - c)^2$? [*Sugerencia*: saque la derivada con respecto a c, iguale a 0 y resuelva.]

b. Utilizando el resultado del inciso (a), ¿cuál de las dos cantidades $\sum(x_i - \bar{x})^2$ y $\sum(x_i - \mu)^2$ será más pequeña que la otra (suponiendo que $\bar{x} \neq \mu$)?

69. a. Sean a y b constantes y sea $y_i = ax_i + b$ con $i = 1, 2, \ldots, n$. ¿Cuáles son las relaciones entre $\bar{x}$ y $\bar{y}$ y entre s_x^2 y s_y^2?

b. Una muestra de temperaturas para iniciar una cierta reacción química dio un promedio muestral (°C) de 87.3 y una desviación estándar muestral de 1.04. ¿Cuáles son el promedio muestral y la desviación estándar medidos en °F? [*Sugerencia*: $F = \frac{9}{5}C + 32$.]

70. El elevado consumo de energía durante el ejercicio continúa después de que termina la sesión de entrenamiento. Debido a que las calorías quemadas por ejercicio contribuyen a la pérdida de peso y tienen otras consecuencias, es importante entender el proceso. El artículo "Effect of Weight Training Exercise and Treadmill Exercise on Post-Exercise Oxygen Consumption" (*Medicine and Science in Sports and Exercise*, 1998: 518–522) reportó los datos adjuntos tomados de un estudio en el cual se midió el consumo de oxígeno (litros) de forma continua durante 30 minutos de cada uno de 15 sujetos tanto después de un entrenamiento con pesas como después de una sesión de ejercicio en una caminadora.

Sujeto	1	2	3	4	5	6	7	
Peso (x)	14.6	14.4	19.5	24.3	16.3	22.1	23.0	
Caminadora (y)	11.3	5.3	9.1	15.2	10.1	19.6	20.8	
Sujeto	8	9	10	11	12	13	14	15
Peso (x)	18.7	19.0	17.0	19.1	19.6	23.2	18.5	15.9
Caminadora (y)	10.3	10.3	2.6	16.6	22.4	23.6	12.6	4.4

a. Construya una gráfica de caja comparativa de las observaciones del ejercicio con pesas y en la caminadora y comente sobre lo que ve.

b. Debido a que estos datos aparecen en pares (x, y), con mediciones de x y y de la misma variable en dos condiciones distintas, es natural enfocarse en las diferencias que existen en ellos: $d_1 = x_1 - y_1, \ldots, d_n = x_n - y_n$. Construya una gráfica de caja de las diferencias muestrales. ¿Qué sugiere la gráfica?

71. La siguiente es una descripción dada por Minitab de los datos de resistencia dados en el ejercicio 13.

```
Variable   N    Mean Median TrMean StDev SE Mean
strength 153 135.39 135.40 135.41  4.59    0.37

Variable    Minimum    Maximum        Q1        Q3
strength     122.20     147.70    132.95    138.25
```

a. Comente sobre cualesquiera características interesantes (los cuartiles y los cuartos son virtualmente idénticos en este caso).

b. Construya una gráfica de caja de los datos basada en los cuartiles y comente sobre lo que ve.

72. Los desórdenes y síntomas de ansiedad con frecuencia pueden ser tratados exitosamente con benzodiazepina. Se sabe que los animales expuestos a estrés exhiben una disminución de la ligadura de receptor de benzodiazepina en la corteza frontal. El artículo "Decreased Benzodiazepine Receptor Binding in Prefrontal Cortex in Combat-Related Posttraumatic Stress Disorder" (*Amer. J. of Psychiatry*, 2000: 1120–1126) describió el primer estudio de ligadura de receptor de benzodiazepina en individuos que sufren de PTSD. Los datos anexos sobre una medición de ligadura a receptor (volumen de distribución ajustado) se leyeron en una gráfica que aparece en el artículo.

PTSD: 10, 20, 25, 28, 31, 35, 37, 38, 38, 39, 39, 42, 46

Saludables: 23, 39, 40, 41, 43, 47, 51, 58, 63, 66, 67, 69, 72

Use varios métodos de este capítulo para describir y resumir los datos.

73. El artículo "Can We Really Walk Straight?" (*Amer. J. of Physical Anthropology*, 1992: 19–27) reportó sobre un experimento en el cual a cada uno de 20 hombres saludables se les pidió que caminaran en línea recta como fuera posible hacia un punto a 60 m de distancia a velocidad normal. Considérense las siguientes observaciones de cadencia (número de pasos por segundo):

.95 .85 .92 .95 .93 .86 1.00 .92 .85 .81
.78 .93 .93 1.05 .93 1.06 1.06 .96 .81 .96

Use los métodos desarrollados en este capítulo para resumir los datos; incluya una interpretación o discusión en los casos en que sea apropiado. [*Nota*: el autor del artículo utilizó un análisis estadístico un tanto complejo para concluir que las personas no pueden caminar en línea recta y sugirió varias explicaciones para esto.]

74. La **moda** de un conjunto de datos numéricos es el valor que ocurre con más frecuencia en el conjunto.

a. Determine la moda de los datos de cadencia dados en el ejercicio 73.

b. Para una muestra categórica, ¿cómo definiría la categoría modal?

75. Se seleccionaron especímenes de tres tipos diferentes de cable y se determinó el límite de fatiga (MPa) de cada espécimen y se obtuvieron los datos adjuntos.

Tipo 1	350	350	350	358	370	370	370	371
	371	372	372	384	391	391	392	
Tipo 2	350	354	359	363	365	368	369	371
	373	374	376	380	383	388	392	
Tipo 3	350	361	362	364	364	365	366	371
	377	377	377	379	380	380	392	

a. Construya una gráfica de caja comparativa y comente sobre las similitudes y diferencias.
b. Construya una gráfica de puntos comparativa (una gráfica de puntos de cada muestra con una escala común). Comente sobre las similitudes y diferencias.
c. ¿Da la gráfica de caja comparativa del inciso (a) una evaluación informativa de similitudes y diferencias? Explique su razonamiento.

76. Las tres medidas de centro introducidas en este capítulo son la media, la mediana y la media recortada. Dos medidas de centro adicionales que de vez en cuando se utilizan son el *rango medio*, el cual es el promedio de las observaciones más pequeñas y más grandes y el *cuarto medio*, el cual es el promedio de los dos cuartos. ¿Cuáles de estas cinco medidas de centro son resistentes a los efectos de los valores apartados y cuáles no? Explique su razonamiento.

77. Los autores del artículo "Predictive Model for Pitting Corrosion in Buried Oil and Gas Pipelines" (Corrosion 2009:332-342) proporcionan los datos en los que basaron sus investigaciones.

a. Considere la muestra siguiente de 61 observaciones de la profundidad de los pozos máxima (mm) de tipos de tubería enterradas en suelo de arcilla limo.

0.41	0.41	0.41	0.41	0.43	0.43	0.43	0.48	0.48
0.58	0.79	0.79	0.81	0.81	0.81	0.91	0.94	0.94
1.02	1.04	1.04	1.17	1.17	1.17	1.17	1.17	1.17
1.17	1.19	1.19	1.27	1.40	1.40	1.59	1.59	1.60
1.68	1.91	1.96	1.96	1.96	2.10	2.21	2.31	2.46
2.49	2.57	2.74	3.10	3.18	3.30	3.58	3.58	4.15
4.75	5.33	7.65	7.70	8.13	10.41	13.44		

Construya una gráfica de tallos y hojas en la cual los dos valores más grandes se muestran en la última fila HI.

b. Use de nuevo el inciso (a), y construya un histograma basado en las ocho clases con 0 como el límite inferior de la primera clase y anchos de clases de .5, .5, .5, 1, 2 y 5, respectivamente.

c. La gráfica de caja de comparativa de Minitab que se observa abajo muestra lotes de la profundidad de los pozos para cuatro tipos diferentes de suelos. Describa sus características importantes.

78. Considere una muestra $x_1, x_2, \ldots, x_n$ y suponga que los valores de $\bar{x}$, s^2 y s han sido calculados.

a. Sea $y_i = x_i - \bar{x}$ con $i = 1, \ldots, n$. ¿Cómo se comparan los valores de s^2 y s de las y_i con los valores correspondientes de las x_i? Explique.
b. Sea $z_i = (x_i - \bar{x})/s$ con $i = 1, \ldots, n$. ¿Cuáles son los valores de la varianza muestral y la desviación estándar muestral de las z_i?

79. Si $\bar{x}_n$ y s_n^2 denotan la media y la varianza de la muestra $x_1, \ldots, x_n$ y si $\bar{x}_{n+1}$ y s_{n+1}^2 denotan estas cantidades cuando se agrega una observación adicional x_{n+1} a la muestra.

a. Demuestre cómo se puede calcular $\bar{x}_{n+1}$ con $\bar{x}_n$ y x_{n+1}.
b. Demuestre que

$$ns_{n+1}^2 = (n-1)s_n^2 + \frac{n}{n+1}(x_{n+1} - \bar{x}_n)^2$$

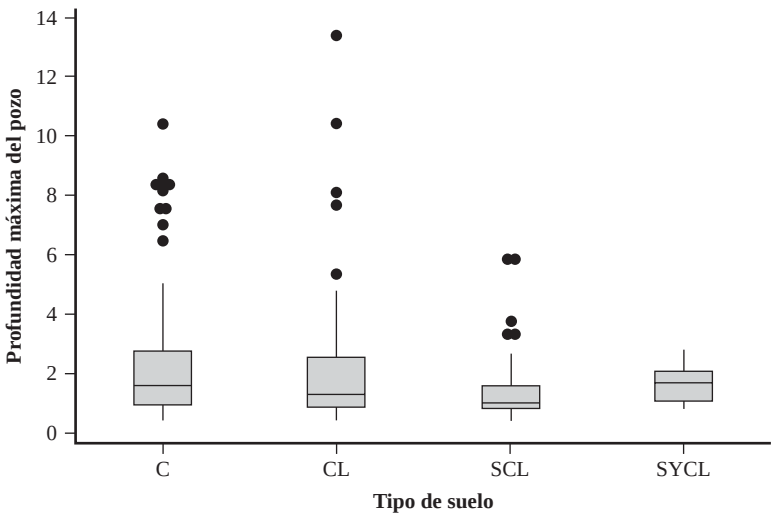
de modo que s_{n+1}^2 pueda ser calculada con x_{n+1}, $\bar{x}_n$ y s_n^2.
c. Suponga que una muestra de 15 torzales de hilo para telas dio por resultado una media muestral del alargamiento del hilo de 12.58 mm y una desviación estándar muestral de .512 mm. Una 16ava torzal resulta en un valor de alargamiento de 11.8. ¿Cuáles son los valores de la media muestral y la desviación estándar muestral de las 16 observaciones de alargamiento?

80. Las distancias de recorrido de rutas de autobuses para cualquier sistema de tránsito particular por lo general varían de una ruta a otra. El artículo "Planning of City Bus Routes" (*J. of the Institution of Engineers*, 1995: 211–215) da la siguiente información sobre las distancias (km) de un sistema particular.

Distancia de recorrido	6−<8	8−<10	10−<12	12−<14	14−<16
Frecuencia	6	23	30	35	32

Distancia de recorrido	16−<18	18−<20	20−<22	22−<24	24−<26
Frecuencia	48	42	40	28	27

Gráfica de caja comparativa para el ejercicio 77



Distancia de recorrido	26−<28	28−<30	30−<35	35−<40	40−<45
Frecuencia	26	14	27	11	2

a. Trace un histograma correspondiente a estas frecuencias.
b. ¿Qué proporción de estas distancias de ruta son menores que 20? ¿Qué proporción de estas rutas tienen distancias de recorrido de por lo menos 30?
c. ¿Aproximadamente cuál es el valor del 90º percentil de la distribución de distancia de recorrido de las rutas?
d. ¿Aproximadamente cuál es la mediana de la distancia de recorrido?

81. Un estudio realizado para investigar la distribución de tiempo de frenado total (tiempo de reacción más tiempo de movimiento de acelerador a freno, en ms) durante condiciones de manejo reales a 60 km/h da la siguiente información sobre la distribución de los tiempos ("A Field Study on Braking Responses During Driving", *Ergonomics*, 1995: 1903–1910):

media = 535 mediana = 500 moda = 500
d. estándar = 96 mínima = 220 máxima = 925
5º percentil = 400 10º percentil = 430
90º percentil = 640 95º percentil = 720

¿Qué puede concluir sobre la forma de un histograma de estos datos? Explique su razonamiento.

82. Los datos muestrales $x_1, x_2, \ldots, x_n$ en ocasiones representan una **serie de tiempo**, donde x_t = el valor observado de una variable de respuesta x en el tiempo t. A menudo la serie observada muestra una gran cantidad de variación aleatoria, lo que dificulta estudiar el comportamiento a plazo más largo. En tales situaciones, es deseable producir una versión suavizada de la serie. Una técnica para hacerlo implica el **suavizamiento exponencial**. Se elige el valor de una constante de suavizamiento α $(0 < \alpha < 1)$. Luego con $\bar{x}_t$ = valor suavizado en el tiempo t, se hace $\bar{x}_1 = x_1$ y para $t = 2, 3, \ldots, n$, $\bar{x}_t = \alpha x_t + (1 - \alpha)\bar{x}_{t-1}$.

a. Considere la siguiente serie de tiempo en la cual x_t = temperatura (°F) del efluente en una planta de tratamiento de aguas negras en el día t: 47, 54, 53, 50, 46, 46, 47, 50, 51, 50, 46, 52, 50, 50. Trace cada x_t contra t en un sistema de coordenadas de dos dimensiones (una gráfica de tiempo-serie). ¿Parece haber algún patrón?

b. Calcule las $\bar{x}_t$ con $\alpha = .1$. Repita con $\alpha = .5$. ¿Qué valor de α da una serie $\bar{x}_t$ más atenuada?
c. Sustituya $\bar{x}_{t-1} = \alpha x_{t-1} + (1 - \alpha)\bar{x}_{t-2}$ en el miembro de la derecha de la expresión para $\bar{x}_t$, acto seguido sustituya $\bar{x}_{t-2}$ en función de x_{t-2} y $\bar{x}_{t-3}$, y así sucesivamente. ¿De cuántos de los valores $x_t, x_{t-1}, \ldots, x_1$ depende $\bar{x}_t$. ¿Qué le sucede al coeficiente de x_{t-k} conforme k se incrementa?
d. Remítase al inciso (c). Si t es grande, ¿qué tan sensible es $\bar{x}_t$ a la inicialización $\bar{x}_1 = x_1$? Explique.

[*Nota*: Una referencia pertinente es el artículo "Simple Statistics for Interpreting Environmental Data", *Water Pollution Control Fed. J.*, 1981: 167–175.]

83. Considere las observaciones numéricas $x_1, \ldots, x_n$. Con frecuencia interesa saber si las x_i están (por lo menos en forma aproximada) simétricamente distribuidas en torno a algún mismo valor. Si n es por lo menos grande de manera moderada, el grado de simetría puede ser valorado con una gráfica de tallo y hojas o un histograma. Sin embargo, si n no es muy grande, las gráficas mencionadas no son informativas en particular. Considere la siguiente alternativa. Que y_1 denote la x_i más pequeña, y_2 la segunda x_i más pequeña, y así sucesivamente. Luego grafique los siguientes pares como puntos en un sistema de coordenadas de dos dimensiones $(y_n - \tilde{x}, \tilde{x} - y_1)$, $(y_{n-1} - \tilde{x}, \tilde{x} - y_2)$, $(y_{n-2} - \tilde{x}, \tilde{x} - y_3), \ldots$ Existen $n/2$ puntos cuando n es par y $(n - 1)/2$ cuando n es impar.

a. ¿Qué apariencia tiene esta gráfica cuando la simetría en los datos es perfecta? ¿Qué apariencia tiene cuando las observaciones se alargan más sobre la mediana que debajo de ella (una larga cola superior)?
b. Los datos adjuntos sobre cantidad de lluvia (acres-pies) producida por 26 nubes bombardeadas se tomaron del artículo "A Bayesian Analysis of a Multiplicative Treatment Effect in Weather Modification" (*Technometrics*, 1975: 161–166). Construya la gráfica y comente sobre el grado de simetría o la naturaleza del alejamiento de la misma.

4.1	7.7	17.5	31.4	32.7	40.6	92.4
115.3	118.3	119.0	129.6	198.6	200.7	242.5
255.0	274.7	274.7	302.8	334.1	430.0	489.1
703.4	978.0	1656.0	1697.8	2745.6		

Bibliografía

Chambers, John, William Cleveland, Beat Kleiner y Paul Tukey, *Graphical Methods for Data Analysis*, Brooks/Cole, Pacific Grove, CA, 1983. Una presentación altamente recomendada de varias metodologías gráficas y pictóricas en estadística.

Cleveland, William, *Visualizing Data*, Hobart Press, Summit, NJ, 1993. Un entretenido recorrido de técnicas pictóricas.

Peck, Roxy y Jay Devore, *Statistics*: *The Exploration and Analysis of Data* (6a. ed.), Thomson Brooks/Cole, Belmont, CA, 2008. Los primeros capítulos hacen un recuento no muy matemático de métodos para describir y resumir datos.

Freedman, David, Robert Pisani y Roger Purves, *Statistics* (4a. ed.), Norton, Nueva York, 2007. Un excelente estudio no muy matemático de razonamiento y metodología estadísticos básicos.

Hoaglin, David, Frederick Mosteller y John Tukey, *Understanding Robust and Exploratory Data Analysis*, Wiley, Nueva York, 1983. Discute por qué y cómo deben ser utilizados los métodos exploratorios; es bueno en lo que se refiere a los detalles de gráficas de tallo y hojas y gráficas de caja.

Moore, David y William Notz, *Statistics: Concepts and Controversies* (7a. ed.), Freeman, San Francisco, 2009. Un libro de pasta blanda extremadamente fácil de leer y ameno que contiene una discusión intuitiva de problemas conectados con experimentos de muestreo y diseñados.

Peck, Roxy y colaboradores. (eds.), *Statistics: A Guide to the Unknown* (4a. ed.), Thomson Brooks/Cole, Belmont, CA, 2006. Contiene muchos artículos cortos no técnicos que describen varias aplicaciones de la estadística.

Verzani, John, *Using R for Introductory Statistics*, Chapman y Hall/CRC, Boca Ratón, FL, 2005. Una introducción muy agradable al paquete de software R.

2 Probabilidad

INTRODUCCIÓN

El término **probabilidad** se refiere al estudio del azar y la incertidumbre en cualquier situación en la cual varios posibles sucesos pueden ocurrir; la disciplina de la probabilidad proporciona métodos de cuantificar las oportunidades y probabilidades asociadas con los varios sucesos. El lenguaje de probabilidad se utiliza constantemente de manera informal tanto en el contexto escrito como en el hablado. Algunos ejemplos incluyen enunciados tales como "es probable que el índice Dow-Jones se incremente al final del año", "existen 50–50 probabilidades de que la persona en posesión de su cargo busque la reelección", "probablemente se ofrecerá por lo menos una sección del curso el próximo año", "las probabilidades favorecen la rápida solución de la huelga" y "se espera que se vendan por lo menos 20,000 boletos para el concierto". En este capítulo se introducen algunos conceptos de probabilidad, se indica cómo pueden ser interpretadas las probabilidades y se demuestra cómo pueden ser aplicadas las reglas de probabilidad para calcular las probabilidades de muchos eventos interesantes. La metodología de probabilidad permite entonces expresar en lenguaje preciso enunciados informales como los antes expresados.

El estudio de la probabilidad como una rama de las matemáticas se remonta a más de 300 años, cuando nace en conexión con preguntas que implicaban juegos de azar. Muchos libros se han ocupado exclusivamente de la probabilidad, pero el objetivo en este caso es abarcar sólo la parte de la materia que tiene más aplicación directa en problemas de inferencia estadística.

2.1 Espacios muestrales y eventos

Un **experimento** es cualquier acción o proceso cuyo resultado está sujeto a la incertidumbre. Aunque la palabra *experimento* en general sugiere una situación de prueba cuidadosamente controlada en un laboratorio, se le utiliza aquí en un sentido mucho más amplio. Por lo tanto experimentos que pueden ser de interés incluyen lanzar al aire una moneda una o varias veces, seleccionar una carta o cartas de un mazo, pesar una hogaza de pan, medir el tiempo de recorrido de la casa al trabajo en una mañana particular, obtener tipos de sangre de un grupo de individuos o medir las resistencias a la compresión de diferentes vigas de acero.

El espacio muestral de un experimento

DEFINICIÓN

El **espacio muestral** de un experimento, denotado por $\mathcal{S}$, es el conjunto de todos los posibles resultados de dicho experimento.

Ejemplo 2.1 El experimento más simple al que se aplica la probabilidad es uno con dos posibles resultados. Tal experimento consiste en examinar un fusible para ver si está defectuoso. El espacio muestral de este experimento se abrevia como $\mathcal{S} = \{N, D\}$, donde N representa no defectuoso, D representa defectuoso, y los paréntesis se utilizan para encerrar los elementos de un conjunto. Otro experimento como ése implicaría lanzar al aire una tachuela y observar si cae punta arriba o punta abajo, con espacio muestral $\mathcal{S} = \{U, D\}$, y otro más consistiría en observar el sexo del siguiente niño nacido en el hospital, con $\mathcal{S} = \{H, M\}$. ■

Ejemplo 2.2 Si se examinan tres fusibles en secuencia y se anota el resultado de cada examen, entonces un resultado del experimento es cualquier secuencia de letras N y D de longitud 3, por lo tanto

$$\mathcal{S} = \{NNN, NND, NDN, NDD, DNN, DND, DDN, DDD\}$$

Si se hubiera lanzado una tachuela tres veces, el espacio muestral se obtendría reemplazando N por U en la expresión $\mathcal{S}$ anterior, y con un cambio de notación similar se obtendría el espacio muestral para el experimento en el cual se observan los sexos de tres niños recién nacidos. ■

Ejemplo 2.3 Dos gasolineras están localizadas en cierta intersección. Cada una dispone de seis bombas de gasolina. Considérese el experimento en el cual se determina el número de bombas en uso a una hora particular del día en cada una de las gasolineras. Un resultado experimental especifica cuántas bombas están en uso en la primera gasolinera y cuántas están en uso en la segunda. Un posible resultado es (2, 2), otro es (4, 1) y otro más es (1, 4). Los 49 resultados en $\mathcal{S}$ se muestran en la tabla adjunta. El espacio muestral del experimento en el cual un dado de seis lados es lanzado dos veces se obtiene eliminando la fila 0 y la columna 0 de la tabla y se obtienen 36 resultados.

		Segunda estación						
		0	**1**	**2**	**3**	**4**	**5**	**6**
Primera estación	**0**	(0, 0)	(0, 1)	(0, 2)	(0, 3)	(0, 4)	(0, 5)	(0, 6)
	1	(1, 0)	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
	2	(2, 0)	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
	3	(3, 0)	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
	4	(4, 0)	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
	5	(5, 0)	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
	6	(6, 0)	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

■

Ejemplo 2.4 Un porcentaje bastante grande de programas C++ escritos en una empresa particular, compilan en la primera ejecución, pero algunos no lo hacen (un *compilador* es un programa que traduce el código fuente, en este caso programas C++, en lenguaje de máquina para que los programas puedan ser ejecutados). Supongamos que un experimento consiste en seleccionar y compilar programas C++ en este lugar uno por uno hasta encontrar uno que compile en la primera ejecución. Se denota un programa que compila en la primera ejecución por S (para el éxito) y uno que no lo hace por F (por error). Aunque puede que no sea muy probable, un posible resultado de este experimento es que los primeros 5 (o 10 o 20 o . . .) sean F y el siguiente sea un S. Es decir, para cualquier entero positivo n, es posible que tenga que examinar n programas antes de ver la primera S. El espacio muestral es $\mathcal{S} = \{S, FS, FFS, FFFS, \ldots\}$, el cual contiene un número infinito de posibles resultados. La misma forma abreviada del espacio muestral es apropiada para un experimento en el cual, a partir de una hora especificada, se anota el sexo de cada infante recién nacido hasta que nazca un varón. ■

Eventos

En el estudio de la probabilidad, interesan no sólo los resultados individuales de $\mathcal{S}$ sino también varias recopilaciones de resultados de $\mathcal{S}$.

DEFINICIÓN

> Un **evento** es cualquier recopilación (subconjunto) de resultados contenidos en el espacio muestral $\mathcal{S}$. Un evento es **simple** si consiste en exactamente un resultado y **compuesto** si consiste en más de un resultado.

Cuando se realiza un experimento, se dice que ocurre un evento particular A si el resultado experimental resultante está contenido en A. En general, ocurrirá exactamente un evento simple, pero muchos eventos compuestos ocurrirán al mismo tiempo.

Ejemplo 2.5 Considérese un experimento en el cual cada uno de tres vehículos que toman una salida de una autopista particular vira a la izquierda (L) o la derecha (R) al final de la rampa de salida. Los ocho posibles resultados que constituyen el espacio muestral son LLL, RLL, LRL, LLR, LRR, RLR, RRL y RRR. Así pues existen ocho eventos simples, entre los cuales están $E_1 = \{LLL\}$ y $E_5 = \{LRR\}$. Algunos eventos compuestos incluyen

$$A = \{RLL, LRL, LLR\} = \text{el evento en que exactamente uno de los tres vehículos vira a la derecha}$$

$$B = \{LLL, RLL, LRL, LLR\} = \text{el evento en que cuando mucho uno de los vehículos vira a la derecha}$$

$$C = \{LLL, RRR\} = \text{el evento en que los tres vehículos viren en la misma dirección}$$

Suponga que cuando se realiza el experimento, el resultado es LLL. Entonces ha ocurrido el evento simple E_1 y por lo tanto también comprende los eventos B y C (pero no A). ■

Ejemplo 2.6 (continuación del ejemplo 2.3) Cuando se observa el número de bombas en uso en cada una de dos gasolineras de seis bombas, existen 49 posibles resultados, por lo que existen 49 eventos simples: $E_1 = \{(0, 0)\}$, $E_2 = \{(0, 1)\}$, . . . , $E_{49} = \{(6, 6)\}$. Ejemplos de eventos compuestos son

$A = \{(0, 0), (1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$ = el evento en que el número de bombas en uso es el mismo en ambas gasolineras

$B = \{(0, 4), (1, 3), (2, 2), (3, 1), (4, 0)\}$ = el evento en que el número total de bombas en uso es cuatro

$C = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ = el evento en que a lo sumo una bomba está en uso en cada gasolinera. ■

Ejemplo 2.7 (continuación del ejemplo 2.4)

El espacio muestral del experimento del programa de compilación contiene un número infinito de resultados, por lo que existe un número infinito de eventos simples. Los eventos compuestos incluyen

$A = \{S, FS, FFS\}$ = el evento en que cuando mucho se examinan tres programas

$E = \{FS, FFFS, FFFFFS, \ldots\}$ = el evento en que se examina un número par de programas ■

Algunas relaciones de la teoría de conjuntos

Un evento es simplemente un conjunto, así que las relaciones y resultados de la teoría elemental de conjuntos pueden ser utilizados para estudiar eventos. Se utilizarán las siguientes operaciones para crear eventos nuevos a partir de eventos dados.

DEFINICIÓN

1. El **complemento** de un evento A, denotado por A', es el conjunto de todos los resultados en $\mathcal{S}$ que no están contenidos en A.
2. La **unión** de dos eventos A y B, denotados por $A \cup B$ y leídos "A *o* B", es el evento que consiste en todos los resultados que están *en* A *o en* B *o en ambos eventos* (de tal suerte que la unión incluya resultados donde tanto A como B ocurren, así también resultados donde ocurre exactamente uno), es decir, todos los resultados en por lo menos uno de los eventos.
3. La **intersección** de dos eventos A y B, denotada por $A \cap B$ y leída "A *y* B", es el evento que consiste en todos los resultados que están *tanto en* A *como en* B.

Ejemplo 2.8 (continuación del ejemplo 2.3)

En el experimento en el cual se observa el número de bombas en uso en una sola gasolinera de seis bombas, sea $A = \{0, 1, 2, 3, 4\}$, $B = \{3, 4, 5, 6\}$ y $C = \{1, 3, 5\}$. Entonces

$$A' = \{5, 6\}, \quad A \cup B = \{0, 1, 2, 3, 4, 5, 6\} = \mathcal{S}, \quad A \cup C = \{0, 1, 2, 3, 4, 5\},$$
$$A \cap B = \{3, 4\}, \quad A \cap C = \{1, 3\}, \quad (A \cap C)' = \{0, 2, 4, 5, 6\}$$ ■

Ejemplo 2.9 (continuación del ejemplo 2.4)

En el experimento de compilación de programas defina A, B y C como

$$A = \{S, FS, FFS\}, \quad B = \{S, FFS, FFFFS\}, \quad C = \{FS, FFFS, FFFFFS, \ldots\}$$

Entonces

$$A' = \{FFFS, FFFFS, FFFFFS, \ldots\}, \; C' = \{S, FFS, FFFFS, \ldots\}$$
$$A \cup B = \{S, FS, FFS, FFFFS\}, \quad A \cap B = \{S, FFS\}$$ ■

En ocasiones A y B no tienen resultados en común, por lo que la intersección de A y B no contiene resultados.

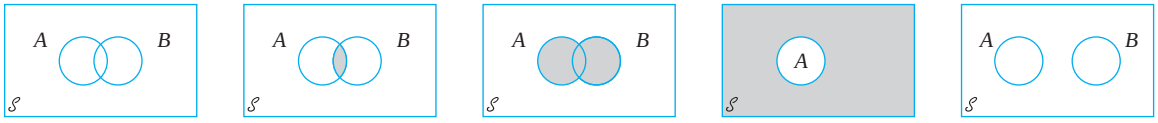
DEFINICIÓN

$\varnothing$ denota el *evento nulo* (el evento sin resultados). Cuando $A \cap B = \varnothing$, se dice que A y B son eventos **mutuamente exclusivos** o **disjuntos.**

Ejemplo 2.10 En una pequeña ciudad hay tres distribuidores de automóviles: un distribuidor GM que vende Chevrolets y Buicks; un distribuidor Ford que vende Fords y Lincolns; y un distribuidor Toyota. Si un experimento consiste en observar la marca del siguiente automóvil vendido, entonces los eventos $A = \{\text{Chevrolet, Buick}\}$ y $B = \{\text{Ford, Lincoln}\}$ son mutuamente exclusivos porque el siguiente automóvil vendido no puede ser tanto un producto GM como un producto Ford (¡a menos que las empresas se fusionen!) ■

Las operaciones de unión e intersección pueden ser ampliadas a más de dos eventos. Para tres eventos cualesquiera A, B y C, el evento $A \cup B \cup C$ es el conjunto de resultados contenidos en por lo menos uno de los tres eventos, mientras que $A \cap B \cap C$ es el conjunto de resultados contenidos en los tres eventos. Se dice que los eventos dados $A_1, A_2, A_3, \ldots$, son mutuamente exclusivos (disjuntos por pares) si ninguno de dos eventos tiene resultados en común.

Con diagramas de Venn se obtiene una representación pictórica de eventos y manipulaciones con eventos. Para construir un diagrama de Venn, se traza un rectángulo cuyo interior representará el espacio muestral $\mathcal{S}$. En tal caso cualquier evento A se representa como el interior de una curva cerrada (a menudo un círculo) contenido en $\mathcal{S}$. La figura 2.1 muestra ejemplos de diagramas de Venn.



(a) Diagrama de Venn de los eventos A y B (b) La región sombreada es $A \cap B$ (c) La región sombreada es $A \cup B$ (d) La región sombreada es A' (e) Eventos mutuamente excluyentes

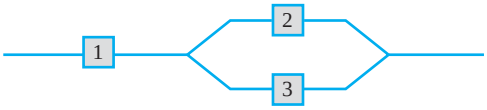
Figura 2.1 Diagramas de Venn

EJERCICIOS Sección 2.1 (1–10)

1. Cuatro universidades, 1, 2, 3 y 4, están participando en un torneo de basquetbol. En la primera ronda, 1 jugará con 2 y 3 jugará con 4. Acto seguido los dos ganadores jugarán por el campeonato y los dos perdedores también jugarán. Un posible resultado puede ser denotado por 1324 (1 derrota a 2 y 3 derrota a 4 en los juegos de la primera ronda, y luego 1 derrota a 3 y 2 derrota a 4).
- **a.** Enumere todos los resultados en $\mathcal{S}$.
- **b.** Que A denote el evento en que 1 gana el torneo. Enumere los resultados en A.
- **c.** Que B denote el evento en que 2 gana el juego de campeonato. Enumere los resultados en B.
- **d.** ¿Cuáles son los resultados en $A \cup B$ y en $A \cap B$? ¿Cuáles son los resultados en A'?

2. Suponga que un vehículo que toma una salida particular de una autopista puede virar a la derecha (R), virar a la izquierda (L) o continuar de frente (S). Observe la dirección de cada uno de tres vehículos sucesivos.
- **a.** Elabore una lista de todos los resultados en el evento A en que los tres vehículos van en la misma dirección.
- **b.** Elabore una lista de todos los resultados en el evento B en que los tres vehículos toman direcciones diferentes.
- **c.** Elabore una lista de todos los resultados en el evento C en que exactamente dos de los tres vehículos dan vuelta a la derecha.
- **d.** Elabore una lista de todos los resultados en el evento D en que dos vehículos van en la misma dirección.
- **e.** Enumere los resultados en D', $C \cup D$ y $C \cap D$.

3. Tres componentes están conectados para formar un sistema como se muestra en el diagrama adjunto. Como los componentes del subsistema 2–3 están conectados en paralelo, dicho subsistema funcionará si por lo menos uno de los dos componentes individuales funciona. Para que todo el sistema funcione, el componente 1 debe funcionar y por lo tanto el subsistema 2–3 debe hacerlo.



El experimento consiste en determinar la condición de cada componente [S (éxito) para un componente que funciona y F (falla) para un componente que no funciona].

a. ¿Qué resultados están contenidos en el evento A en que exactamente dos de los tres componentes funcionan?
b. ¿Qué resultados están contenidos en el evento B en que por lo menos dos de los componentes funcionan?
c. ¿Qué resultados están contenidos en el evento C en que el sistema funciona?
d. Ponga en lista los resultados en C', $A \cup C$, $A \cap C$, $B \cup C$ y $B \cap C$.

4. Cada una de una muestra de cuatro hipotecas residenciales está clasificada como tasa fija (F) o tasa variable (V).
a. ¿Cuáles son los 16 resultados en $\mathcal{S}$?
b. ¿Qué resultados están en el evento en que exactamente tres de las hipotecas seleccionadas son de tasa fija?
c. ¿Qué resultados están en el evento en que las cuatro hipotecas son del mismo tipo?
d. ¿Qué resultados están en el evento en que a lo sumo una de las cuatro es una hipoteca de tasa variable?
e. ¿Cuál es la unión de eventos en los incisos (c) y (d), y cuál es la intersección de estos dos eventos?
f. ¿Cuáles son la unión e intersección de los dos eventos en los incisos (b) y (c)?

5. Una familia compuesta de tres personas, A, B y C, acude a una clínica médica que siempre tiene disponible un doctor en cada una de las estaciones 1, 2 y 3. Durante cierta semana, cada miembro de la familia visita la clínica una vez y es asignado al azar a una estación. El experimento consiste en registrar la estación para cada miembro. Un resultado es (1, 2, 1) para A a la estación 1, B a la estación 2 y C a la estación 1.
a. Elabore una lista de los 27 resultados en el espacio muestral.
b. Elabore una lista de todos los resultados en el evento en que los tres miembros van a la misma estación.
c. Haga una lista de los resultados en el evento en el que todos los miembros van a diferentes estaciones.
d. Elabore una lista de los resultados en el evento en que ninguno va a la estación 2.

6. La biblioteca de una universidad dispone de cinco ejemplares de un cierto texto en reserva. Dos ejemplares (1 y 2) son primeras impresiones y los otros tres (3, 4 y 5) son segundas impresiones. Un estudiante examina estos libros en orden aleatorio y se detiene sólo cuando una segunda impresión ha sido seleccionada. Un posible resultado es 5 y otro 213.
a. Ponga en lista los resultados en $\mathcal{S}$.
b. Sea A el evento en que exactamente un libro debe ser examinado. ¿Qué resultados están en A?
c. Sea B el evento en que el libro 5 es seleccionado. ¿Qué resultados están en B?
d. Sea C el evento en que el libro 1 no es examinado. ¿Qué resultados están en C?

7. Un departamento académico acaba de votar en secreto para elegir un jefe de departamento. La urna contiene cuatro boletas con votos para el candidato A y tres con votos para el candidato B. Suponga que estas boletas se sacan de la urna una por una.
a. Ponga en lista todos los posibles resultados.
b. Suponga que mantiene un conteo continuo de las boletas retiradas de la urna. ¿Para qué resultados A se mantiene adelante de B durante todo el conteo?

8. Una firma constructora de ingeniería en la actualidad está trabajando en plantas eléctricas en tres sitios diferentes. Que A_i denote el evento en que la planta localizada en el sitio i se complete alrededor de la fecha contratada. Use las operaciones de unión, intersección y complementación para describir cada uno de los siguientes eventos en función de A_1, A_2 y A_3, trace un diagrama de Venn y sombree la región que corresponde a cada uno.
a. Por lo menos una planta se completa alrededor de la fecha contratada.
b. Todas las plantas se completan alrededor de la fecha contratada.
c. Sólo la planta localizada en el sitio 1 se completa alrededor de la fecha contratada.
d. Exactamente una planta se completa alrededor de la fecha contratada.
e. La planta localizada en el sitio 1 o las otras dos plantas se completan alrededor de la fecha contratada.

9. Use diagramas de Venn para verificar las dos siguientes relaciones para los eventos A y B (éstas se conocen como leyes de De Morgan):
a. $(A \cup B)' = A' \cap B'$
b. $(A \cap B)' = A' \cup B'$

[*Sugerencia:* en cada inciso dibuje un diagrama que corresponda al lado derecho y otro al izquierdo.]

10. **a.** En el ejemplo 2.10, identifique tres eventos que sean mutuamente exclusivos.
b. Suponga que no hay resultado común a los tres eventos A, B y C. ¿Son estos tres eventos mutuamente exclusivos por necesidad? Si su respuesta es sí, explique por qué; si su respuesta es no, dé un contraejemplo valiéndose del experimento del ejemplo 2.10.

2.2 Axiomas, interpretaciones, y propiedades de la probabilidad

Dados un experimento y un espacio muestral $\mathcal{S}$, *el objetivo de la probabilidad es asignar a cada evento A un número P(A), llamado la probabilidad del evento A, el cual dará una medida precisa de la oportunidad de que A ocurrirá*. Para garantizar que las asignaciones

serán consistentes con las nociones intuitivas de la probabilidad, todas las asignaciones deberán satisfacer los siguientes axiomas (propiedades básicas) de probabilidad.

AXIOMA 1 Para cualquier evento A, $P(A) \geq 0$.

AXIOMA 2 $P(\mathcal{S}) = 1$.

AXIOMA 3 Si $A_1, A_2, A_3, \ldots$ es un conjunto de eventos disjuntos, entonces

$$P(A_1 \cup A_2 \cup A_3 \cup \cdots) = \sum_{i=1}^{\infty} P(A_i)$$

Se podría preguntar por qué el tercer axioma no contiene ninguna referencia a un conjunto *finito* de eventos disjuntos. Es porque la propiedad correspondiente para un conjunto finito puede ser derivada de los tres axiomas. Se pretende que la lista de axiomas sea tan corta como sea posible y que no contenga alguna propiedad que pueda ser derivada de las demás que aparecen en la lista. El axioma 1 refleja la noción intuitiva de que la probabilidad de que ocurra A deberá ser no negativa. El espacio muestral es por definición el evento que debe ocurrir cuando se realiza el experimento ($\mathcal{S}$ contiene todos los posibles resultados), así que el axioma 2 dice que la máxima probabilidad posible de 1 está asignada a $\mathcal{S}$. El tercer axioma formaliza la idea que si se desea la probabilidad de que por lo menos uno de varios eventos ocurrirá y dado que dos eventos no pueden ocurrir al mismo tiempo, entonces la probabilidad de que por lo menos uno ocurra es la suma de las probabilidades de los eventos individuales.

PROPOSICIÓN $P(\varnothing) = 0$ donde $\varnothing$ es el evento nulo (el evento que no contiene resultados en absoluto). Esto a su vez implica que la propiedad contenida en el axioma 3 es válida para un conjunto *finito* de eventos disjuntos.

Comprobación Primero considérese el conjunto infinito $A_1 = \varnothing, A_2 = \varnothing, A_3 = \varnothing, \ldots$ Como $\varnothing \cap \varnothing = \varnothing$, los eventos en este conjunto están disjuntos y $\bigcup A_i = \varnothing$. El tercer axioma da entonces

$$P(\varnothing) = \sum P(\varnothing)$$

Esto puede suceder sólo si $P(\varnothing) = 0$.

Ahora supóngase que $A_1, A_2, \ldots, A_k$ son eventos disjuntos y anéxese a éstos el conjunto infinito $A_{k+1} = \varnothing, A_{k+2} = \varnothing, A_{k+3} = \varnothing, \ldots$. Si de nuevo se invoca el tercer axioma,

$$P\left(\bigcup_{i=1}^{k} A_i\right) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^{k} P(A_i)$$

como se deseaba. ■

Ejemplo 2.11 Considere lanzar una tachuela al aire. Cuando se detiene en el suelo, su punta estará hacia arriba (el resultado U) o hacia abajo (el resultado D). El espacio muestral de este evento es por consiguiente $\mathcal{S} = \{U, D\}$. Los axiomas especifican $P(\mathcal{S}) = 1$, por lo que la asignación de probabilidad se completará determinando $P(U)$ y $P(D)$. Como U y D están disjuntos y su unión es $\mathcal{S}$, la siguiente proposición implica que

$$1 = P(\mathcal{S}) = P(U) + P(D)$$

Se deduce que $P(D) = 1 - P(U)$. Una posible asignación de probabilidades es $P(U) = .5$, $P(D) = .5$, mientras que otra posible asignación es $P(U) = .75$, $P(D) = .25$. De hecho, si p representa cualquier número fijo entre 0 y 1, $P(U) = p$, $P(D) = 1 - p$ es una asignación compatible con los axiomas. ■

Ejemplo 2.12 Considere probar las baterías que salen de la línea de ensamble una por una hasta que se encuentra una con el voltaje dentro de los límites prescritos. Los eventos simples son $E_1 = \{S\}$, $E_2 = \{FS\}$, $E_3 = \{FFS\}$, $E_4 = \{FFFS\}$, ... Suponga que la probabilidad de que cualquier batería resulte satisfactoria es de .99. Entonces se puede demostrar que $P(E_1) = .99$, $P(E_2) = (.01)(.99)$, $P(E_3) = (.01)^2(.99)$, . . . es una asignación de probabilidades a los eventos simples que satisface los axiomas. En particular, como los E_i son disjuntos y $\mathcal{S} = E_1 \cup E_2 \cup E_3 \cup \ldots$, debe ser el caso que

$$\begin{aligned} 1 = P(\mathcal{S}) &= P(E_1) + P(E_2) + P(E_3) + \cdots \\ &= .99[1 + .01 + (.01)^2 + (.01)^3 + \cdots] \end{aligned}$$

Aquí se utilizó la fórmula para la suma de una serie geométrica:

$$a + ar + ar^2 + ar^3 + \cdots = \frac{a}{1 - r}$$

Sin embargo, otra asignación de probabilidad legítima (de acuerdo con los axiomas) del mismo tipo "geométrico" se obtiene reemplazando .99 por cualquier otro número p entre 0 y 1 (y .01 por $1 - p$). ■

Interpretación de probabilidad

Los ejemplos 2.11 y 2.12 muestran que los axiomas no determinan por completo una asignación de probabilidades a eventos. Los axiomas sirven sólo para excluir las asignaciones incompatibles con las nociones intuitivas de probabilidad. En el experimento de lanzar al aire tachuelas del ejemplo 2.11, se sugirieron dos asignaciones particulares. La asignación apropiada o correcta depende de la naturaleza de la tachuela y también de la interpretación de probabilidad. La interpretación que más frecuentemente se utiliza y es más fácil de entender está basada en la noción de frecuencias relativas.

Considérese un experimento que pueda ser realizado repetidamente de una manera idéntica e independiente, y sea A un evento que consiste en un conjunto fijo de resultados del experimento. Ejemplos simples de experimentos repetibles incluyen el lanzamiento al aire de tachuelas y dados previamente discutido. Si el experimento se realiza n veces, en algunas de las réplicas el evento A ocurrirá (el resultado estará en el conjunto A) y en otras, A no ocurrirá. Denote con $n(A)$ el número de réplicas en las cuales A sí ocurre. Entonces la relación $n(A)/n$ se conoce como la *frecuencia relativa* de ocurrencia del evento A en la secuencia de n réplicas.

Por ejemplo, sea A el evento de que un paquete enviado dentro del estado de California para entrega en un segundo día en realidad llega en un día. Los resultados de enviar 10 paquetes de este tipo (las primeras 10 repeticiones) son los siguientes:

Paquete #	1	2	3	4	5	6	7	8	9	10
¿A ocurre?	N	S	S	S	N	N	S	S	N	N
Frecuencia relativa de A	0	.5	.667	.75	.6	.5	.571	.625	.556	.5

La figura 2.2(a) muestra cómo la frecuencia relativa $n(A)/n$ fluctúa sustancialmente en el curso de las primeras 50 repeticiones. Pero como el número de repeticiones sigue aumentando, la figura 2.2(b) ilustra cómo la frecuencia relativa se estabiliza.

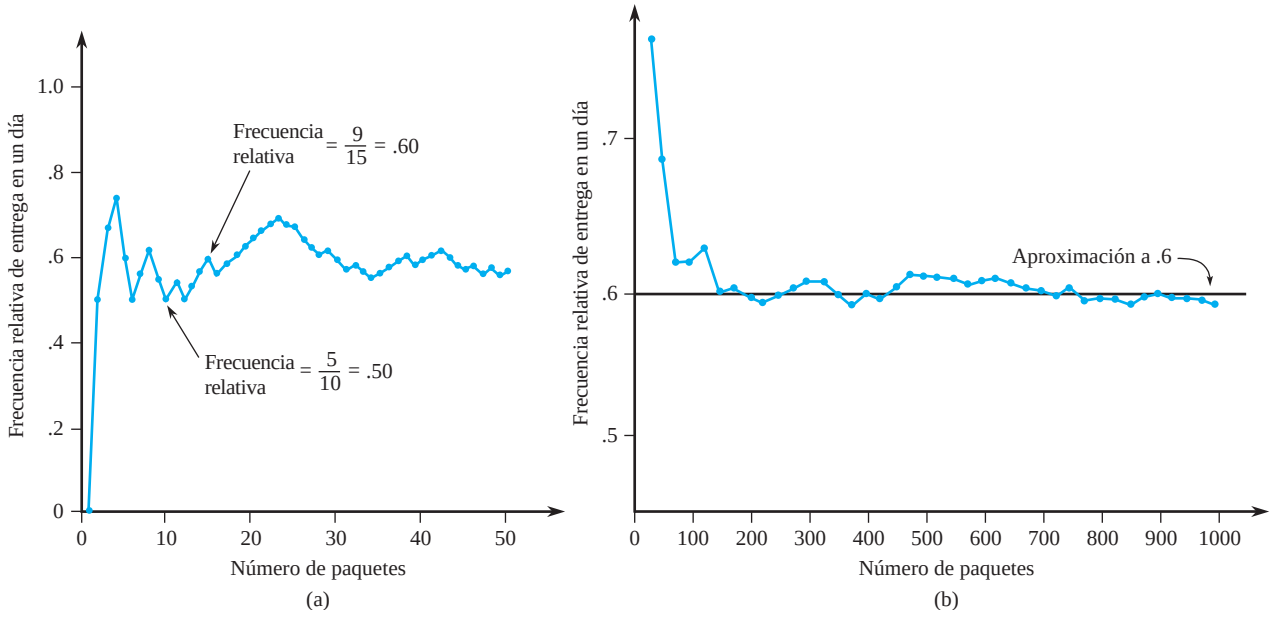


Figura 2.2 Comportamiento de la frecuencia relativa (a) fluctuación inicial (b) estabilización a largo plazo.

En términos más generales, la evidencia empírica, con base en los resultados de muchos experimentos repetibles, indica que cualquier frecuencia relativa de este tipo se estabilizará conforme el número de repeticiones n aumenta. Es decir, como n se hace arbitrariamente grande, $n(A)/n$ se aproxima a un valor límite al que se denomina *límite* (o *largo plazo*) *de la frecuencia relativa* del evento A. La *interpretación objetiva de probabilidad* identifica esta frecuencia límite en relación con $P(A)$. Supóngase que las probabilidades se asignan a los acontecimientos de acuerdo con su límite de frecuencias relativas. A continuación, una afirmación como "la probabilidad de un paquete se entregue en un día de envío es .6" significa que de un gran número de paquetes enviados por correo, aproximadamente el 60% llegará en un día. Del mismo modo, si B es el evento de que un tipo particular de aparato necesitará servicio mientras la garantía es válida, entonces $P(B) = .1$ se interpreta en el sentido de que en el largo plazo, el 10% de estos aparatos necesitará un servicio de garantía. Esto no quiere decir que exactamente uno de cada 10 necesitará servicio o que exactamente 10 de cada 100 necesitarán servicio, ya que 10 y 100 no son a largo plazo.

Se dice que esta interpretación de frecuencia relativa de probabilidad es objetiva porque se apoya en una propiedad del experimento y no en algún individuo particular interesado en el experimento. Por ejemplo, dos observadores diferentes de una secuencia de lanzamiento de una moneda deberán utilizar la misma asignación de probabilidad puesto que los observadores no tienen nada que ver con la frecuencia relativa límite. En la práctica, la interpretación no es tan objetiva como pudiera parecer, puesto que la frecuencia relativa límite de un evento no será conocida. Por tanto, se tendrán que asignar probabilidades con base en creencias sobre la frecuencia relativa límite de eventos en estudio. Afortunadamente, existen muchos experimentos para los cuales habrá consenso con respecto a asignaciones de probabilidad. Cuando se habla de una moneda imparcial, significa $P(A) = P(S) = .5$ y un dado imparcial es uno para el cual las frecuencias relativas limitativas de los seis resultados son $\frac{1}{6}$, lo que sugiere las asignaciones de probabilidad $P(\{1\}) = \cdots = P(\{6\}) = \frac{1}{6}$.

Como la interpretación objetiva de probabilidad está basada en la noción de frecuencia limitativa, su aplicabilidad está limitada a situaciones experimentales repetibles. No obstante, el lenguaje de probabilidad a menudo se utiliza en conexión con situaciones

que son inherentemente irrepetibles. Algunos ejemplos incluyen: "Las probabilidades de un tratado de paz son buenas"; "es probable que el contrato le sea otorgado a nuestra compañía"; y "como su mejor mariscal de campo está lesionado, espero que no anoten más de 10 puntos contra nosotros". En tales situaciones se desearía, como antes, asignar probabilidades numéricas a varios resultados y eventos (p. ej., la probabilidad es .9 de que obtendremos el contrato). Por consiguiente se debe adoptar una interpretación alternativa de estas probabilidades. Como diferentes observadores pueden tener información y opiniones previas con respecto a tales situaciones experimentales, las asignaciones de probabilidad ahora pueden diferir de un individuo a otro. Las interpretaciones en tales situaciones se conocen por lo tanto como *subjetivas*. El libro de Robert Winkler citado en las referencias del capítulo da un recuento muy fácil de leer de varias interpretaciones subjetivas.

Más propiedades de probabilidad

PROPOSICIÓN

Para cualquier evento A, $P(A) + P(A') = 1$, a partir de la cual $P(A) = 1 - P(A')$.

Comprobación En el axioma 3, sea $k = 2$, $A_1 = A$ y $A_2 = A'$. Como por la definición de A', $A \cup A' = \mathcal{S}$ en tanto A y A' sean eventos disjuntos, $1 = P(\mathcal{S}) = P(A \cup A') = P(A) + P(A')$. ■

Esta proposición es sorprendentemente útil porque se presentan muchas situaciones en las cuales $P(A')$ es más fácil de obtener mediante métodos directos que $P(A)$.

Ejemplo 2.13 Considere un sistema de cinco componentes idénticos conectados en serie, como se ilustra en la figura 2.3.

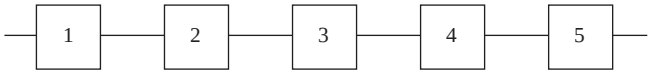


Figura 2.3 Sistema de cinco componentes conectados en serie

Denote un componente que falla por F y uno que no lo hace por S. Sea A el evento en que el *sistema* falla. Para que ocurra A, por lo menos uno de los componentes individuales debe fallar. Los resultados en A incluyen $SSFSS$ (1, 2, 4 y 5 funcionarán, pero 3 no), $FFSSS$, y así sucesivamente. Existen de hecho 31 resultados diferentes en A. Sin embargo, A', el evento en que el sistema funciona, consiste en el resultado único $SSSSS$. En la sección 2.5 se verá que si 90% de todos estos componentes no fallan y diferentes componentes fallan independientemente uno de otro, entonces $P(A') = P(SSSSS) = .9^5 = .59$. Así pues, $P(A) = 1 - .59 = .41$; por lo tanto, entre un gran número de sistemas como ése, aproximadamente 41% fallarán. ■

En general, la proposición anterior es útil cuando el evento de interés puede ser expresado como "por lo menos . . . ," puesto que en ese caso puede ser más fácil trabajar con el complemento "menos que . . ." (en algunos problemas es más fácil trabajar con "más que . . ." que con "cuando mucho . . ."). Cuando se tenga dificultad al calcular $P(A)$ directamente, habrá que pensar en determinar $P(A')$.

PROPOSICIÓN

Para cualquier evento A, $P(A) \leq 1$.

Esto se debe a que $1 = P(A) + P(A') \geq P(A)$ puesto que $P(A') \geq 0$.

Cuando los eventos A y B son mutuamente exclusivos, $P(A \cup B) = P(A) + P(B)$. Para eventos que no son mutuamente exclusivos, la adición de $P(A)$ y $P(B)$ da por resultado un "doble conteo" de los resultados en la intersección. El siguiente resultado muestra cómo corregir esto.

PROPOSICIÓN

Para dos eventos cualesquiera A y B,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Comprobación Obsérvese primero que $A \cup B$ puede ser descompuesto en dos eventos *excluyentes*, A y $B \cap A'$; la última es la parte de B que queda afuera de A (ver figura 2.4). Además, B por sí mismo es la unión de los dos eventos excluyentes $A \cap B$ y $A' \cap B$, por lo tanto $P(B) = P(A \cap B) + P(A' \cap B)$. Así

$$\begin{aligned} P(A \cup B) = P(A) + P(B \cap A') &= P(A) + [P(B) - P(A \cap B)] \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

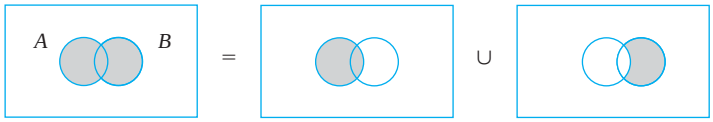


Figura 2.4 Representación de A ∪ B como la unión de dos eventos excluyentes ■

Ejemplo 2.14 En cierto suburbio residencial, 60% de las familias se suscriben al servicio de internet de la compañía de televisión por cable, 80% lo hacen al servicio de televisión de esa compañía y 50% de todas las familias a ambos servicios. Si se elige una familia al azar, ¿cuál es la probabilidad de que contrate por lo menos uno de estos dos servicios de la empresa, y cuál es la probabilidad de contratar exactamente uno de estos servicios de la empresa?

Con $A = \{$se suscribe al servicio de internet$\}$ y $B = \{$se suscribe al servicio de televisión por cable$\}$, la información dada implica que $P(A) = .6$, $P(B) = .8$ y $P(A \cap B) = .5$. La proposición precedente ahora lleva a

$P($se suscribe a por lo menos uno de los dos servicios$)$

$$= P(A \cup B) = P(A) + P(B) - P(A \cap B) = .6 + .8 - .5 = .9$$

El evento de que una familia se suscribe sólo al servicio de televisión por cable se escribe como $A' \cap B$ [(no internet) y televisión]. Ahora la figura 2.4 implica que

$$.9 = P(A \cup B) = P(A) + P(A' \cap B) = .6 + P(A' \cap B)$$

a partir de la cual $P(A' \cap B) = .3$. Asimismo, $P(A \cap B') = P(A \cup B) - P(B) = .1$. Todo esto se ilustra en la figura 2.5, donde se ve que

$$P(\text{exactamente uno}) = P(A \cap B') + P(A' \cap B) = .1 + .3 = .4$$

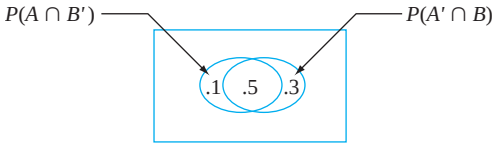


Figura 2.5 Probabilidades para el ejemplo 2.14 ■

La probabilidad de una unión de más de dos eventos se calcula en forma análoga.

Para tres eventos cualesquiera A, B y C,

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Esto se puede verificar examinando un diagrama de Venn de $A \cup B \cup C$, el cual se muestra en la figura 2.6. Cuando $P(A)$, $P(B)$ y $P(C)$ se agregan, ciertas intersecciones se cuentan dos veces, por lo que deben ser restadas, pero esto hace que $P(A \cap B \cap C)$ se reste una vez con demasiada frecuencia.

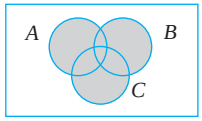


Figura 2.6 $A \cup B \cup C$

Determinación sistemática de probabilidades

Considérese un espacio muestral que es o finito o "contablemente infinito" (lo segundo significa que los resultados pueden ser puestos en lista en una secuencia infinita, por lo que existe un primer resultado, un segundo, un tercero, y así sucesivamente, por ejemplo, el escenario de prueba de baterías del ejemplo 2.12). Denote con $E_1, E_2, E_3, \ldots$ los eventos simples correspondientes, cada uno compuesto de un solo resultado. Una estrategia sensible para el cálculo de probabilidad es determinar primero cada probabilidad del evento simple, con el requerimiento de que $\sum P(E_i) = 1$. Entonces la probabilidad de cualquier evento compuesto A se calcula agregando los $P(E_i)$ para todos los E_i que existen en A:

$$P(A) = \sum_{\text{todos los } E_i \text{ en } A} P(E_i)$$

Ejemplo 2.15 Durante las horas no pico el tren que viaja entre los suburbios y la ciudad utiliza cinco carros. Suponga que existe el doble de probabilidades de que un usuario seleccione el carro intermedio (#3) que cualquier carro adyacente (#2 o #4) y el doble de probabilidades de que seleccione cualquier carro adyacente que cualquier carro extremo (#1 o #5). Sea $p_i = P(\text{carro } i \text{ es seleccionado}) = P(E_i)$. Entonces se tiene $p_3 = 2p_2 = 2p_4$ y $p_2 = 2p_1 = 2p_5 = p_4$. Esto da

$$1 = \sum P(E_i) = p_1 + 2p_1 + 4p_1 + 2p_1 + p_1 = 10p_1$$

es decir, $p_1 = p_5 = .1$, $p_2 = p_4 = .2$, $p_3 = .4$. La probabilidad de que uno de los tres carros intermedios se seleccione (un evento compuesto) es entonces $p_2 + p_3 + p_4 = .8$. ■

Resultados igualmente probables

En muchos experimentos compuestos de N resultados, es razonable asignar probabilidades iguales a todos los N eventos simples. Esto incluye ejemplos tan obvios como lanzar al aire una moneda o un dado imparciales una o dos veces (o cualquier número fijo de veces) o seleccionar una o varias cartas de un mazo bien barajado de 52 cartas. Con $p = P(E_i)$ por cada i,

$$1 = \sum_{i=1}^{N} P(E_i) = \sum_{i=1}^{N} p = p \cdot N \text{ por lo tanto } p = \frac{1}{N}$$

Es decir, si existen N resultados igualmente probables, la probabilidad de cada uno es $1/N$.

Ahora considérese un evento A, con $N(A)$ como el número de resultados contenidos en A. Entonces

$$P(A) = \sum_{E_i \text{ en } A} P(E_i) = \sum_{E_i \text{ en } A} \frac{1}{N} = \frac{N(A)}{N}$$

Por lo tanto, cuando los resultados son igualmente probables, el cálculo de probabilidades se reduce a contar: determinar tanto el número de resultados $N(A)$ en A como el número de resultados N en $\mathcal{S}$ y formar su cociente.

Ejemplo 2.16 Usted tiene seis libros de misterios sin leer y seis de ciencia ficción sin leer en su biblioteca. Los tres primeros de cada tipo son de tapa dura y los tres últimos son de bolsillo. Considere la posibilidad de seleccionar al azar uno de los seis misterios y luego seleccionar al azar uno de los seis libros de ciencia ficción para tomar unas vacaciones en Acapulco (después de todo, necesita algo para leer en la playa). Numera los de misterio 1, 2, . . . , 6, y hace lo mismo con los libros de ciencia ficción. A continuación, cada resultado es un par de números, tales como (4, 1) y hay $N = 36$ resultados posibles (para una representación visual de esta situación, consulte la tabla en el ejemplo 2.3 y elimine la primera fila y columna). Con la selección al azar como se ha descrito, los 36 resultados son igualmente probables. Nueve de estos resultados son tales que los libros seleccionados son libros de bolsillo (aquellos en la esquina inferior derecha de la tabla de referencia): (4, 4), (4, 5), . . . , (6, 6). Así que la probabilidad del evento A de que ambos libros seleccionados sean libros de bolsillo es

$$P(A) = \frac{N(A)}{N} = \frac{9}{36} = .25$$ ■

EJERCICIOS Sección 2.2 (11–28)

11. Una compañía de fondos de inversión mutua ofrece a sus clientes varios fondos diferentes: un fondo de mercado de dinero, tres fondos de bonos (a corto, intermedio y largo plazos), dos fondos de acciones (de moderado y alto riesgo) y un fondo balanceado. Entre los clientes que poseen acciones en un solo fondo, los porcentajes de clientes en los diferentes fondos son como sigue:

Mercado de dinero	20%	Acciones de alto riesgo	18%
Bonos a corto plazo	15%	Acciones de riesgo moderado	25%
Bonos a plazo intermedio	10%	Balanceadas	7%
Bonos a largo plazo	5%		

Se selecciona al azar un cliente que posee acciones en sólo un fondo.

a. ¿Cuál es la probabilidad de que el individuo seleccionado posea acciones en el fondo balanceado?

b. ¿Cuál es la probabilidad de que el individuo posea acciones en un fondo de bonos?

c. ¿Cuál es la probabilidad de que el individuo seleccionado no posea acciones en un fondo de acciones?

12. Considere seleccionar al azar un estudiante en cierta universidad y que A denote el evento en que el individuo seleccionado tenga una tarjeta de crédito Visa y que B sea el evento análogo para la tarjeta MasterCard. Suponga que $P(A) = .5$, $P(B) = .4$, y $P(A \cap B) = .25$.

a. Calcule la probabilidad de que el individuo seleccionado tenga por lo menos uno de los dos tipos de tarjetas (es decir, la probabilidad del evento $A \cup B$).

b. ¿Cuál es la probabilidad de que el individuo seleccionado no tenga ningún tipo de tarjeta?

c. Describa, en función de A y B, el evento de que el estudiante seleccionado tenga una tarjeta Visa pero no una MasterCard y luego calcule la probabilidad de este evento.

13. Una firma consultora de computación presentó propuestas en tres proyectos. Sea $A_i = \{\text{proyecto otorgado } i\}$, con $i = 1, 2, 3$ y suponga que $P(A_1) = .22$, $P(A_2) = .25$, $P(A_3) = .28$, $P(A_1 \cap A_2) = .11$, $P(A_1 \cap A_3) = .05$, $P(A_2 \cap A_3) = .07$, $P(A_1 \cap A_2 \cap A_3) = .01$. Exprese en palabras cada uno de los siguientes eventos y calcule la probabilidad de cada uno:

a. $A_1 \cup A_2$

b. $A_1' \cap A_2'$ [*Sugerencia:* $(A_1 \cup A_2)' = A_1' \cap A_2'$]

c. $A_1 \cup A_2 \cup A_3$ **d.** $A_1' \cap A_2' \cap A_3'$

e. $A_1' \cap A_2' \cap A_3$ **f.** $(A_1' \cap A_2') \cup A_3$

14. Suponga que el 55% de los adultos consumen regularmente café, el 45% consumen regularmente refrescos con gas y el 70% consumen regularmente al menos uno de estos dos productos.

a. ¿Cuál es la probabilidad de que un adulto al azar regularmente consuma café y soda?
b. ¿Cuál es la probabilidad de que un adulto al azar no consuma regularmente al menos uno de estos dos productos?

15. Considere el tipo de secadora de ropa (de gas o eléctrica) adquirida por cada uno de cinco clientes diferentes en cierta tienda.
a. Si la probabilidad de que a lo sumo uno de éstos adquiera una secadora eléctrica es .428, ¿cuál es la probabilidad de que por lo menos dos adquieran una secadora eléctrica?
b. Si P(los cinco compran una secadora de gas) = .116 y P(los cinco compran una secadora eléctrica) = .005, ¿cuál es la probabilidad de que por lo menos se adquiera una secadora de cada tipo?

16. A un individuo se le presentan tres vasos diferentes de refresco de cola, designados C, D y P. Se le pide que pruebe los tres y que los ponga en lista en orden de preferencia. Suponga que se sirvió el mismo refresco de cola en los tres vasos.
a. ¿Cuáles son los eventos simples en este evento de clasificación y qué probabilidad le asignaría a cada uno?
b. ¿Cuál es la probabilidad de que C obtenga el primer lugar?
c. ¿Cuál es la probabilidad de que C obtenga el primer lugar y D el último?

17. Denote con A el evento en que la siguiente solicitud de asesoría de un consultor de software estadístico tenga que ver con el paquete SPSS y que B denote el evento en que la siguiente solicitud de ayuda tiene que ver con SAS. Suponga que $P(A) = .30$ y $P(B) = .50$.
a. ¿Por qué no es el caso que $P(A) + P(B) = 1$?
b. Calcule $P(A')$.
c. Calcule $P(A \cup B)$.
d. Calcule $P(A' \cap B')$.

18. Una caja contiene seis focos de 40 W, cinco de 60 W y cuatro de 75 W. Si los focos se eligen uno por uno en orden aleatorio, ¿cuál es la probabilidad de que por lo menos dos focos deban ser seleccionados para obtener uno de 75 W?

19. La inspección visual humana de uniones soldadas en un circuito impreso puede ser muy subjetiva. Una parte del problema se deriva de los numerosos tipos de defectos de soldadura (p. ej., almohadilla seca, visibilidad en escuadra, picaduras) e incluso el grado al cual una unión posee uno o más de estos defectos. Por consiguiente, incluso inspectores altamente entrenados pueden discrepar en cuanto a la disposición particular de una unión particular. En un lote de 10,000 uniones, el inspector A encontró 724 defectuosas, el inspector B 751 y 1159 de las uniones fueron consideradas defectuosas por cuando menos uno de los inspectores. Suponga que se selecciona una de las 10,000 uniones al azar.
a. ¿Cuál es la probabilidad de que la unión seleccionada no sea juzgada defectuosa por ninguno de los dos inspectores?
b. ¿Cuál es la probabilidad de que la unión seleccionada sea juzgada defectuosa por el inspector B pero no por el inspector A?

20. Cierta fábrica utiliza tres turnos diferentes. Durante el año pasado, ocurrieron 200 accidentes en la fábrica. Algunos de ellos pueden ser atribuidos por lo menos en parte a condiciones de trabajo inseguras mientras que las otras no se relacionan con las condiciones de trabajo. La tabla adjunta da el porcentaje de accidentes que ocurren en cada tipo de categoría de accidente–turno.

		Condiciones inseguras	No vinculados a las condiciones
Turno	***Diurno***	10%	35%
	Mixto	8%	20%
	Nocturno	5%	22%

Suponga que uno de los 200 reportes de accidente se selecciona al azar de un archivo de reportes y que el turno y el tipo de accidente se determinan.
a. ¿Cuáles son los eventos simples?
b. ¿Cuál es la probabilidad de que el accidente seleccionado se atribuya a condiciones inseguras?
c. ¿Cuál es la probabilidad de que el accidente seleccionado no haya ocurrido en el turno de día?

21. Una compañía de seguros ofrece cuatro diferentes niveles de deducible, ninguno, bajo, medio y alto, para sus tenedores de pólizas de propietario de casa y tres diferentes niveles, bajo, medio y alto, para sus tenedores de pólizas de automóviles. La tabla adjunta da proporciones de las varias categorías de tenedores de pólizas que tienen ambos tipos de seguro. Por ejemplo, la proporción de individuos con deducible bajo de casa y deducible bajo de automóvil es .06 (6% de todos los individuos).

	Propietarios de viviendas			
Auto	**N**	**B**	**M**	**A**
B	.04	.06	.05	.03
M	.07	.10	.20	.10
A	.02	.03	.15	.15

Suponga que se elige al azar un individuo que posee ambos tipos de pólizas.
a. ¿Cuál es la probabilidad de que el individuo tenga un deducible de auto medio y un deducible de casa alto?
b. ¿Cuál es la probabilidad de que el individuo tenga un deducible de casa bajo y un deducible de auto bajo?
c. ¿Cuál es la probabilidad de que el individuo se encuentre en la misma categoría de deducibles de casa y auto?
d. Basado en su respuesta en el inciso (c), ¿cuál es la probabilidad de que las dos categorías sean diferentes?
e. ¿Cuál es la probabilidad de que el individuo tenga por lo menos un nivel deducible bajo?
f. Utilizando la respuesta del inciso (e), ¿cuál es la probabilidad de que ningún nivel deducible sea bajo?

22. La ruta utilizada por un automovilista para trasladarse a su trabajo contiene dos intersecciones con señales de tránsito. La probabilidad de que tenga que detenerse en la primera señal es .4, la probabilidad análoga para la segunda señal es .5 y la probabilidad de que tenga que detenerse en por lo menos una de las dos señales es .6. ¿Cuál es la probabilidad de que tenga que detenerse
a. ¿En ambas señales?

b. ¿En la primera señal pero no en la segunda?
c. ¿En exactamente una señal?

23. Las computadoras de seis miembros del cuerpo de profesores en cierto departamento tienen que ser reemplazadas. Dos de ellos seleccionaron computadoras portátiles y los otros cuatro escogieron computadoras de escritorio. Suponga que sólo dos de las configuraciones pueden ser realizadas en un día particular y las dos computadoras que van a ser configuradas se seleccionan al azar de entre las seis (lo que implica 15 resultados igualmente probables; si las computadoras se numeran 1, 2, . . . , 6, entonces un resultado se compone de las computadoras 1 y 2, otro de las computadoras 1 y 3, y así sucesivamente).
a. ¿Cuál es la probabilidad de que las dos configuraciones seleccionadas sean computadoras portátiles?
b. ¿Cuál es la probabilidad de que ambas configuraciones seleccionadas sean computadoras de escritorio?
c. ¿Cuál es la probabilidad de que por lo menos una configuración seleccionada sea una computadora de escritorio?
d. ¿Cuál es la probabilidad de que por lo menos una computadora de cada tipo sea elegida para configurarla?

24. Demuestre que si un evento A está contenido en otro evento B (es decir, A es un subconjunto de B), entonces $P(A) \leq P(B)$. [*Sugerencia*: disjuntos A y B, A y $B \cap A'$ son eventos para tales y $B = A \cup (B \cap A')$, como se ve en el diagrama de Venn.] Para los eventos A y B, ¿qué implica esto sobre la relación entre $P(A \cap B)$, $P(A)$ y $P(A \cup B)$?

25. Las tres opciones más populares en un tipo de automóvil nuevo son un GPS (sistema de posicionamiento global)(A) un quemacocos (B) y una transmisión automática (C). Si 40% de todos los compradores solicitan A, 55% solicitan B, 70% solicitan C, 63% solicitan A o B, 77% solicitan A o C, 80% solicitan B o C y 85% solicitan A o B o C, calcule las probabilidades de los siguientes eventos. [*Sugerencia*: "A o B" es el evento en que por lo menos una de las dos opciones es solicitada; trate de trazar un diagrama de Venn y rotule todas las regiones.]
a. El siguiente comprador solicitará por lo menos una de las tres opciones.
b. El siguiente comprador no seleccionará ninguna de las tres opciones.
c. El siguiente comprador solicitará sólo una transmisión automática y ninguna de las otras dos opciones.
d. El siguiente comprador seleccionará exactamente una de estas tres opciones.

26. Un sistema puede experimentar tres tipos diferentes de defectos. Sea A_i ($i = 1, 2, 3$) el evento en que el sistema tiene un defecto de tipo i. Suponga que

$$P(A_1) = .12 \quad P(A_2) = .07 \quad P(A_3) = .05$$
$$P(A_1 \cup A_2) = .13 \quad P(A_1 \cup A_3) = .14$$
$$P(A_2 \cup A_3) = .10 \quad P(A_1 \cap A_2 \cap A_3) = .01$$

a. ¿Cuál es la probabilidad de que el sistema no tenga un defecto de tipo 1?
b. ¿Cuál es la probabilidad de que el sistema tenga tanto defectos de tipo 1 como de tipo 2?
c. ¿Cuál es la probabilidad de que el sistema tenga tanto defectos de tipo 1 como de tipo 2 pero no de tipo 3?
d. ¿Cuál es la probabilidad de que el sistema tenga a lo sumo dos de estos defectos?

27. Un departamento académico con cinco miembros del cuerpo de profesores, Anderson, Box, Cox, Cramer y Fisher, debe seleccionar a dos de ellos para que participen en un comité de revisión de personal. Como el trabajo requerirá mucho tiempo, ninguno está ansioso de participar, por lo que se decidió que el representante será elegido introduciendo los nombres en cinco trozos de papel dentro una caja, revolviéndolos y seleccionando dos.
a. ¿Cuál es la probabilidad de que tanto Anderson como Box sean seleccionados? [*Sugerencia*: nombre los resultados igualmente probables.]
b. ¿Cuál es la probabilidad de que por lo menos uno de los dos miembros cuyo nombre comienza con C sea seleccionado?
c. Si los cinco miembros del cuerpo de profesores han dado clase durante 3, 6, 7, 10 y 14 años, respectivamente, en la universidad, ¿cuál es la probabilidad de que los dos representantes seleccionados tengan por lo menos 15 años de experiencia académica en la universidad?

28. En el ejercicio 5, suponga que cualquier individuo que entre a la clínica tiene las mismas probabilidades de ser asignado a cualquiera de las tres estaciones independientemente de a dónde hayan sido asignados otros individuos.
¿Cuál es la probabilidad de que
a. los tres miembros de una familia sean asignados a la misma estación?
b. a lo sumo dos miembros de la familia sean asignados a la misma estación?
c. cada miembro de la familia sea asignado a una estación diferente?

2.3 Técnicas de conteo

Cuando los diversos resultados de un experimento son igualmente probables (la misma probabilidad es asignada a cada evento simple), la tarea de calcular probabilidades se reduce a contar. Sea N el número de resultados en un espacio muestral y $N(A)$ el número de resultados contenidos en un evento A,

$$P(A) = \frac{N(A)}{N} \tag{2.1}$$

Si una lista de resultados es fácil de obtener y N es pequeña, entonces N y $N(A)$ pueden ser determinadas sin utilizar ningún principio de conteo.

Existen, sin embargo, muchos experimentos en los cuales el esfuerzo implicado al elaborar la lista es prohibitivo porque N es bastante grande. Explotando algunas reglas de conteo generales, es posible calcular probabilidades de la forma (2.1) sin una lista de resultados. Estas reglas también son útiles en muchos problemas que implican resultados que no son igualmente probables. Se utilizarán varias de las reglas desarrolladas aquí al estudiar distribuciones de probabilidad en el siguiente capítulo.

La regla de producto para pares ordenados

La primera regla de conteo se aplica a cualquier situación en la cual un conjunto (evento) se compone de pares ordenados de objetos y se desea contar el número de pares. Por par ordenado se quiere decir que si O_1 y O_2 son objetos, entonces el par (O_1, O_2) es diferente del par (O_2, O_1). Por ejemplo, si un individuo selecciona una línea aérea para un viaje de Los Ángeles a Chicago y (después de realizar transacciones de negocios en Chicago) una segunda para continuar a Nueva York, una posibilidad es (American, United), otra es (United, American) y otra más es (United, United).

PROPOSICIÓN

Si el primer elemento u objeto de un par ordenado puede ser seleccionado de n_1 maneras, y por cada una de estas n_1 maneras el segundo elemento del par puede ser seleccionado de n_2 maneras, entonces el número de pares es n_1n_2.

Una interpretación alternativa consiste en llevar a cabo una operación que consta de dos etapas. Si la primera etapa se puede realizar en cualquiera de n_1 maneras y para cada una hay n_2 formas de realizar la segunda etapa, entonces, n_1n_2 es el número de maneras de llevar a cabo las dos etapas en la secuencia.

Ejemplo 2.17 El propietario de una casa que va a llevar a cabo una remodelación requiere los servicios tanto de un contratista de fontanería como de un contratista de electricidad. Si existen 12 contratistas de fontanería y 9 contratistas electricistas disponibles en el área, ¿de cuántas maneras pueden ser elegidos los contratistas? Si $P_1, \ldots, P_{12}$ son los fontaneros y $Q_1, \ldots, Q_9$ son los electricistas, entonces se desea el número de pares de la forma (P_i, Q_j). Con $n_1 = 12$ y $n_2 = 9$, la regla de producto da $N = (12)(9) = 108$ formas posibles de seleccionar los dos tipos de contratistas. ■

En el ejemplo 2.17, la selección del segundo elemento del par no dependió de qué primer elemento ocurrió o fue elegido. En tanto exista el mismo número de opciones del segundo elemento por cada primer elemento, la regla de producto es válida incluso cuando el conjunto de posibles segundos elementos depende del primer elemento.

Ejemplo 2.18 Una familia se acaba de cambiar a una nueva ciudad y requiere los servicios tanto de un obstetra como de un pediatra. Existen dos clínicas médicas fácilmente accesibles y cada una tiene dos obstetras y tres pediatras. La familia obtendrá los máximos beneficios del seguro de salud si se une a una clínica y selecciona ambos doctores de dicha clínica. ¿De cuántas maneras se puede hacer esto? Denote los obstetras por O_1, O_2, O_3 y O_4 y los pediatras por $P_1, \ldots, P_6$. Entonces se desea el número de pares (O_i, P_j) para los cuales O_i y P_j están asociados con la misma clínica. Como existen cuatro obstetras, $n_1 = 4$, y por cada uno existen tres opciones de pediatras, por lo tanto $n_2 = 3$. Aplicando la regla de producto se obtienen $N = n_1n_2 = 12$ posibles opciones. ■

En muchos problemas de conteo y probabilidad se puede utilizar una configuración conocida como **diagrama de árbol** para representar pictóricamente todas las posibilidades. El diagrama de árbol asociado con el ejemplo 2.18 aparece en la figura 2.7. Partiendo de un punto localizado en el lado izquierdo del diagrama, por cada posible primer elemento de

un par emana un segmento de línea recta hacia la derecha. Cada una de estas líneas se conoce como rama de primera generación. Ahora para cualquier rama de primera generación se construye otro segmento de línea que emana de la punta de la rama por cada posible opción de un segundo elemento del par. Cada segmento de línea es una rama de segunda generación. Como existen cuatro obstetras, existen cuatro ramas de primera generación y tres pediatras por cada obstetra, resultan tres ramas de segunda generación que emanan de cada rama de primera generación.

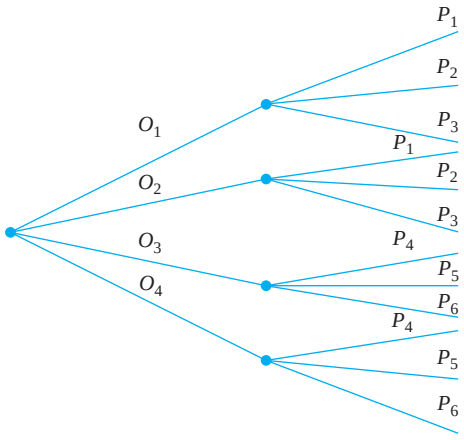


Figura 2.7 Diagrama de árbol para el ejemplo 2.18

Generalizando, supóngase que existen n_1 ramas de primera generación y por cada rama de primera generación existen n_2 ramas de segunda generación. El número total de ramas de segunda generación es entonces n_1n_2. Como el extremo de cada rama de segunda generación corresponde a exactamente un posible par (la selección de un primer elemento y luego de un segundo nos sitúa en el extremo de exactamente una rama de segunda generación), existen n_1n_2 pares, lo que verifica la regla de producto.

La construcción de un diagrama de árbol no depende de tener el mismo número de ramas de segunda generación que emanen de cada rama de primera generación. Si la segunda clínica tuviera cuatro pediatras, entonces habría sólo tres ramas emanando de dos de las ramas de primera generación y cuatro emanando de cada una de las otras dos ramas de primera generación. Un diagrama de árbol puede ser utilizado por lo tanto para representar pictóricamente experimentos diferentes de aquellos a los que se aplica la regla de producto.

Una regla de producto más general

Si se lanza al aire un dado de seis lados cinco veces en sucesión en lugar de sólo dos veces, entonces cada posible resultado es un conjunto ordenado de cinco números tal como (1, 3, 1, 2, 4) o (6, 5, 2, 2, 2). Un conjunto ordenado de k objetos recibirá el nombre de ***k*-tupla** (por tanto, un par es una 2-tupla y una terna es una 3-tupla). Cada resultado del experimento de lanzamiento al aire de un dado es entonces una 5-tupla.

Regla de producto para *k*-tuplas

Supóngase que un conjunto se compone de conjuntos ordenados de k elementos (k-tuplas) y que existen n_1 posibles opciones para el primer elemento; por cada opción del primer elemento, existen n_2 posibles opciones del segundo elemento; . . . ; por cada posible opción de los primeros $k - 1$ elementos, existen n_k opciones del elemento k-ésimo. Existen entonces $n_1n_2 \cdot \cdots \cdot n_k$ posibles k-tuplas.

Una interpretación alternativa consiste en llevar a cabo una operación en k etapas. Si la primera etapa se puede realizar en cualquiera de n_1 maneras, y para cada una de tales maneras hay n_2 formas de realizar la segunda etapa, y para cada forma de llevar a cabo las dos primeras etapas hay n_3 formas de realizar la tercera fase, y así sucesivamente, entonces $n_1n_2 \cdot \cdots \cdot n_k$ es el número de formas para llevar a cabo toda la k-etapa de operación en secuencia. Esta regla más general también se puede visualizar con un diagrama de árbol. Para el caso $k = 3$, sólo se tiene que añadir un número adecuado de una 3ª generación en las ramas de la punta de cada rama de 2ª generación. Si, por ejemplo, una ciudad universitaria tiene cuatro lugares de pizza, un complejo de cine con seis pantallas y tres lugares para ir a bailar, entonces habría cuatro ramas de 1ª generación, seis ramas de 2ª generación que emanan de la punta de cada rama de 1ª generación, y tres ramas de 3ª generación que abren cada rama de 2ª generación. Cada posible 3-tupla corresponde a la punta de una rama de 3ª generación.

Ejemplo 2.19 (continuación del ejemplo 2.17)

Suponga que el trabajo de remodelación de la casa implica adquirir primero varios utensilios de cocina. Se adquirirán en la misma tienda y hay cinco tiendas en el área. Con las tiendas denotadas por $D_1, \ldots, D_5$, existen $N = n_1n_2n_3 = (5)(12)(9) = 540$ 3-tuplas de la forma (D_i, P_j, Q_k), así que existen 540 formas de elegir primero una tienda, luego un contratista de fontanería y finalmente un contratista electricista. ■

Ejemplo 2.20 (continuación del ejemplo 2.18)

Si cada clínica tiene dos especialistas en medicina interna y dos médicos generales, existen $n_1n_2n_3n_4 = (4)(3)(3)(2) = 72$ formas de seleccionar un doctor de cada tipo, de tal suerte que todos los doctores practiquen en la misma clínica. ■

Permutaciones y combinaciones

Considérese un grupo de n individuos u objetos distintos ("distintos" significa que existe alguna característica que diferencia a cualquier individuo u objeto de cualquier otro). ¿Cuántas maneras existen de seleccionar un subconjunto de tamaño k del grupo? Por ejemplo, si un equipo de ligas pequeñas tiene 15 jugadores registrados, ¿cuántas maneras existen de seleccionar 9 jugadores para una alineación inicial? O si una librería universitaria vende diez computadoras portátiles diferentes, pero tiene espacio para mostrar sólo tres de ellas, de cuántas maneras puede elegir las tres?

Una respuesta a la pregunta general que se acaba de plantear requiere distinguir entre dos casos. En algunas situaciones, tal como el escenario del beisbol, el orden de la selección es importante. Por ejemplo, con Ángela como lanzador y Ben como receptor se obtiene una alineación diferente de aquella con Ángela como receptor y Ben como lanzador. A menudo, sin embargo, el orden no es importante y a nadie le interesa qué individuos u objetos sean seleccionados, como sería el caso en el escenario de selección de las computadoras portátiles.

DEFINICIÓN

Un subconjunto ordenado se llama **permutación**. El número de permutaciones de tamaño k que se puede formar con los n individuos u objetos en un grupo será denotado por $P_{k,n}$. Un subconjunto no ordenado se llama **combinación.** Una forma de denotar el número de combinaciones es $C_{k,n}$, pero en su lugar se utilizará una notación que es bastante común en libros de probabilidad: $\binom{n}{k}$, que se lee "de n se elige k".

El número de permutaciones se determina utilizando la primera regla de conteo para k-tuplas. Supóngase, por ejemplo, que un colegio de ingeniería tiene siete departamentos, denotados por a, b, c, d, e, f y g. Cada departamento tiene un representante en el consejo de estudiantes del colegio. De estos siete representantes, uno tiene que ser elegido como

presidente, otro como vicepresidente y un tercero como secretario. ¿Cuántas maneras de seleccionar los tres oficiales existen? Es decir, ¿cuántas permutaciones de tamaño 3 pueden ser formadas con los 7 representantes? Para responder esta pregunta, habrá que pensar en formar una terna (3-tupla) en el cual el primer elemento es el presidente, el segundo es el vicepresidente y el tercero es el secretario. Una terna es (a, g, b), otro es (b, g, a) y otro más es (d, f, b). Ahora bien, el presidente puede ser seleccionado en cualesquiera de $n_1 = 7$ formas. Por cada forma de seleccionar el presidente, existen $n_2 = 6$ formas de seleccionar el vicepresidente y por consiguiente $7 \times 6 = 42$ (pares de presidente, vicepresidente). Por último, por cada forma de seleccionar un presidente y vicepresidente, existen $n_3 = 5$ formas de seleccionar el secretario. Esto da

$$P_{3,7} = (7)(6)(5) = 210$$

como el número de permutaciones de tamaño 3 que se pueden formar con 7 individuos distintos. Una representación de diagrama de árbol mostraría tres generaciones de ramas.

La expresión para $P_{3,7}$ puede ser reescrita con la ayuda de *notación factorial*. Recuérdese que 7! (se lee "factorial de 7") es una notación compacta para el producto descendente de enteros (7)(6)(5)(4)(3)(2)(1). Más generalmente, para cualquier entero positivo m, $m! = m(m-1)(m-2)\cdot\cdots\cdot(2)(1)$. Esto da $1! = 1$, y también se define $0! = 1$. Entonces

$$P_{3,7} = (7)(6)(5) = \frac{(7)(6)(5)(4!)}{(4!)} = \frac{7!}{4!}$$

Más generalmente,

$$P_{k,n} = n(n-1)(n-2)\cdot\cdots\cdot(n-(k-2))(n-(k-1))$$

Multiplicando y dividiendo ésta por $(n-k)!$ se obtiene una expresión compacta para el número de permutaciones.

PROPOSICIÓN

$$P_{k,n} = \frac{n!}{(n-k)!}$$

Ejemplo 2.21 Existen diez asistentes de profesor disponibles para calificar exámenes en un curso de cálculo en una gran universidad. El primer examen se compone de cuatro preguntas y el profesor desea seleccionar un asistente diferente para calificar cada pregunta (sólo un asistente por pregunta). ¿De cuántas maneras se pueden elegir los asistentes para calificar? En este caso n = tamaño del grupo = 10 y k = tamaño del subconjunto = 4. El número de permutaciones es

$$P_{4,10} = \frac{10!}{(10-4)!} = \frac{10!}{6!} = 10(9)(8)(7) = 5040$$

Es decir, el profesor podría aplicar 5040 exámenes diferentes de cuatro preguntas sin utilizar la misma asignación de calificadores a las preguntas, ¡tiempo en el cual todos los asistentes seguramente habrán terminado sus programas de licenciatura! ■

Considérense ahora las combinaciones (es decir, subconjuntos no ordenados). De nuevo habrá que remitirse al escenario de consejo estudiantil, y supóngase que tres de los siete representantes tienen que ser seleccionados para que asistan a una convención estatal. El orden de selección no es importante; lo que importa es cuáles tres son seleccionados. Así que se busca $\binom{7}{3}$, el número de combinaciones de 3 que se pueden formar con los

7 individuos. Considérese por un momento las combinaciones *a*, *c*, *g*. Estos tres individuos pueden ser ordenados en 3! = 6 formas para producir el número de permutaciones:

$$a, c, g \quad a, g, c \quad c, a, g \quad c, g, a \quad g, a, c \quad g, c, a$$

De manera similar, hay 3! = 6 maneras para ordenar la combinación *b*, *c*, *e* para producir permutaciones, y de hecho hay 3! modos de ordenar cualquier combinación particular de tamaño 3 para producir permutaciones. Esto implica la siguiente relación entre el número de combinaciones y el número de permutaciones:

$$P_{3,7} = (3!)\cdot\binom{7}{3} \Rightarrow \binom{7}{3} = \frac{P_{3,7}}{3!} = \frac{7!}{(3!)(4!)} = \frac{(7)(6)(5)}{(3)(2)(1)} = 35$$

No sería difícil poner en lista las 35 combinaciones, pero no hay necesidad de hacerlo si sólo interesa cuántas son. Obsérvese que el número de 210 permutaciones excede por mucho el número de combinaciones; ¡el primero es más grande que el segundo por un factor de 3! puesto que así es como cada combinación puede ser ordenada.

Generalizando la línea de razonamiento anterior se obtiene una relación simple entre el número de permutaciones y el número de combinaciones que produce una expresión concisa para la última cantidad.

PROPOSICIÓN

$$\binom{n}{k} = \frac{P_{k,n}}{k!} = \frac{n!}{k!(n-k)!}$$

Nótese que $\binom{n}{n} = 1$ y $\binom{n}{0} = 1$ puesto que hay sólo una forma de seleccionar un conjunto de *n* (todos) elementos o de ningún elemento, y $\binom{n}{1} = n$ puesto que existen *n* subconjuntos de tamaño 1.

Ejemplo 2.22 Una lista de reproducción de iPod contiene 100 canciones, de las cuales 10 son de los Beatles. Supongamos que la función de reproducción aleatoria se utiliza para reproducir las canciones en orden aleatorio (la aleatoriedad del proceso de barajar es investigada en "¿Su iPod *realmente* reproduce favoritos?" (*The Amer. Statistician*, 2009: 263–268). ¿Cuál es la probabilidad de que la primera canción de los Beatles escuchada sea la quinta canción reproducida?

Para que este evento ocurra, debe ser el caso de que las primeras cuatro canciones que se reproducen no sean canciones de los Beatles (NB) y que la quinta canción sea de los Beatles (B). El número de maneras de seleccionar las primeras cinco canciones es de 100(99)(98)(97)(96). El número de maneras de seleccionar estas cinco canciones para que las cuatro primeras sean NB y la siguiente sea B es de 90(89)(88)(87)(10). La suposición aleatoria implica que cualquier conjunto particular de 5 canciones de entre las 100 tiene la misma probabilidad de ser seleccionado como los primeros cinco reproducidos al igual que cualquier otro conjunto de cinco canciones; cada resultado es igualmente probable. Por lo tanto, la probabilidad deseada es la relación entre el número de resultados para que el evento de interés ocurra con el número de resultados posibles:

$$P(1^{a}\text{ B es la }5^{a}\text{ canción reproducida}) = \frac{90\cdot 89\cdot 88\cdot 87\cdot 10}{100\cdot 99\cdot 98\cdot 97\cdot 96} = \frac{P_{4,90}\cdot(10)}{P_{5,100}} = .0679$$

A continuación está una línea alternativa de razonamiento que implica combinaciones. En lugar de centrarse en la selección de sólo las primeras cinco canciones, piense en reproducir las 100 canciones en orden aleatorio. El número de formas de elegir 10 de estas canciones que sean B (sin tener en cuenta el orden en que se reprodujeron) es $\binom{100}{10}$. Ahora bien, si elegimos 9 de las últimas 95 canciones que sean B, que se puede hacer de $\binom{95}{9}$ maneras, deja cuatro NB y una B para las primeras cinco canciones. Sólo hay una forma

más de estas cinco para empezar con cuatro NB y luego seguir con una B (recordemos que estamos considerando subconjuntos *desordenados*). Por lo tanto

$$P(1^{a}\text{ B es la }5^{a}\text{ canción reproducida}) = \frac{\binom{95}{9}}{\binom{100}{10}}$$

Es fácil verificar que esta última expresión es, de hecho, idéntica a la primera expresión para la probabilidad deseada, por lo que el resultado numérico es de nuevo 0.0679.

La probabilidad de que una de las primeras cinco canciones que se reproducen sea una canción de los Beatles es

$$P(1^{a}\text{ B es la }1^{a}\text{ o }2^{a}\text{ o }3^{a}\text{ o }4^{a}\text{ o }5^{a}\text{ canción reproducida})$$
$$= \frac{\binom{99}{9}}{\binom{100}{10}} + \frac{\binom{98}{9}}{\binom{100}{10}} + \frac{\binom{97}{9}}{\binom{100}{10}} + \frac{\binom{96}{9}}{\binom{100}{10}} + \frac{\binom{95}{9}}{\binom{100}{10}} = .4162$$

Por tanto, es bastante probable que la canción de los Beatles sea una de las primeras cinco canciones reproducidas. Esta "coincidencia" no es tan sorprendente como podría parecer. ■

Ejemplo 2.23 El almacén de una universidad recibió 25 impresoras, de las cuales 10 son impresoras láser y 15 son modelos de inyección de tinta. Si 6 de estas 25 se seleccionan al azar para que las revise un técnico particular, ¿cuál es la probabilidad de que exactamente 3 de las seleccionadas sean impresoras láser (de modo que las otras 3 sean de inyección de tinta)?

Sea $D_3 = \{\text{exactamente 3 de las 6 seleccionadas son impresoras de inyección de tinta}\}$. Suponiendo que cualquier conjunto particular de 6 impresoras es tan probable de ser elegido como cualquier otro conjunto de 6, se tienen resultados igualmente probables, por lo tanto $P(D_3) = N(D_3)/N$, donde N es el número de formas de elegir 6 impresoras de entre las 25 y $N(D_3)$ es el número de formas de elegir 3 impresoras láser y 3 de inyección de tinta. Por lo tanto $N = \binom{25}{6}$. Para obtener $N(D_3)$, primero se piensa en elegir 3 de las 15 impresoras de inyección de tinta y luego 3 de las impresoras láser. Existen $\binom{15}{3}$ formas de elegir las 3 impresoras de inyección de tinta y $\binom{10}{3}$ formas de elegir las 3 impresoras láser; $N(D_3)$ es ahora el producto de estos dos números (visualícese un diagrama de árbol; en realidad aquí se está utilizando el argumento de la regla de producto), por lo tanto

$$P(D_3) = \frac{N(D_3)}{N} = \frac{\binom{15}{3}\binom{10}{3}}{\binom{25}{6}} = \frac{\frac{15!}{3!12!}\cdot\frac{10!}{3!7!}}{\frac{25!}{6!19!}} = .3083$$

Sea $D_4 = \{\text{exactamente 4 de las 6 impresoras seleccionadas son impresoras de inyección de tinta}\}$ y defínanse D_5 y D_6 del mismo modo. Entonces la probabilidad de seleccionar por lo menos 3 impresoras de inyección de tinta es

$$P(D_3 \cup D_4 \cup D_5 \cup D_6) = P(D_3) + P(D_4) + P(D_5) + P(D_6)$$
$$= \frac{\binom{15}{3}\binom{10}{3}}{\binom{25}{6}} + \frac{\binom{15}{4}\binom{10}{2}}{\binom{25}{6}} + \frac{\binom{15}{5}\binom{10}{1}}{\binom{25}{6}} + \frac{\binom{15}{6}\binom{10}{0}}{\binom{25}{6}} = .8530$$

■

EJERCICIOS Sección 2.3 (29–44)

29. Con fecha de abril de 2006, aproximadamente 50 millones de nombres de dominio web.com fueron registrados (p. ej., yahoo.com).
- **a.** ¿Cuántos nombres de dominio compuestos de exactamente dos letras en sucesión pueden ser formados? ¿Cuántos nombres de dominio de dos letras existen si como caracteres se permiten dígitos y letras? [*Nota*: una longitud de carácter de tres o más ahora es obligatoria.]
- **b.** ¿Cuántos nombres de dominio existen compuestos de tres letras en secuencia? ¿Cuántos de esta longitud existen si se permiten letras o dígitos? [*Nota*: en la actualidad todos están utilizados.]
- **c.** Responda las preguntas hechas en (b) para secuencias de cuatro caracteres.
- **d.** Con fecha de abril de 2006, 97,786 de las secuencias de cuatro caracteres utilizando o letras o dígitos aún no habían sido reclamadas. Si se elige un nombre de cuatro caracteres al azar, ¿cuál es la probabilidad de que ya tenga dueño?

30. Un amigo mío va a ofrecer una fiesta. Sus existencias actuales de vino incluye 8 botellas de zinfandel, 10 de merlot y 12 de cabernet (él sólo bebe vino tinto), todos de diferentes fábricas vinícolas.
- **a.** Si desea servir 3 botellas de zinfandel y el orden de servicio es importante, ¿cuántas formas existen de hacerlo?
- **b.** Si 6 botellas de vino tienen que ser seleccionadas al azar de las 30 para servirse, ¿cuántas formas existen de hacerlo?
- **c.** Si se seleccionan al azar 6 botellas, ¿cuántas formas existen de obtener dos botellas de cada variedad?
- **d.** Si se seleccionan 6 botellas al azar, ¿cuál es la probabilidad de que el resultado sea dos botellas de cada variedad?
- **e.** Si se eligen 6 botellas al azar, ¿cuál es la probabilidad de que todas ellas sean de la misma variedad?

31. **a.** Beethoven escribió 9 sinfonías y Mozart 27 conciertos para piano. Si el locutor de una estación de radio de una universidad desea transmitir primero una sinfonía de Beethoven y luego un concierto de Mozart, ¿de cuántas maneras puede hacerlo?
- **b.** El gerente de la estación decide que en cada noche sucesiva (7 días a la semana), se tocará una sinfonía de Beethoven, seguida por un concierto para piano de Mozart, seguido por un cuarteto de cuerdas de Schubert (de los cuales existen 15). ¿Durante aproximadamente cuántos años se podría continuar con esta política antes de que exactamente el mismo programa se repitiera?

32. Una tienda de equipos de sonido está ofreciendo un precio especial en un juego completo de componentes (receptor, reproductor de discos compactos, altavoces, tornamesa). Al comprador se le ofrece una opción de fabricante por cada componente.

Receptor: Kenwood, Onkyo, Pioneer, Sony, Sherwood
Reproductor de discos compactos: Onkyo, Pioneer, Sony, Technics
Altavoces: Boston, Infinity, Polk
Tornamesa: Onkyo, Sony, Teac, Technics

Un tablero de distribución en la tienda permite al cliente conectar cualquier selección de componentes (compuesta de uno de cada tipo). Use las reglas de producto para responder las siguientes preguntas.
- **a.** ¿De cuántas maneras puede ser seleccionado un componente de cada tipo?
- **b.** ¿De cuántas maneras pueden ser seleccionados los componentes si tanto el receptor como el reproductor de discos compactos tiene que ser Sony?
- **c.** ¿De cuántas maneras pueden ser seleccionados los componentes si ninguno tiene que ser Sony?
- **d.** ¿De cuántas maneras se puede hacer una selección si por lo menos se tiene que incluir un componente Sony?
- **e.** Si alguien mueve los interruptores en el tablero de distribución completamente al azar, ¿cuál es la probabilidad de que el sistema seleccionado contenga por lo menos un componente Sony? ¿Exactamente un componente Sony?

33. De nuevo considere el equipo de ligas pequeñas que tiene 15 jugadores en su plantel.
- **a.** ¿Cuántas formas existen de seleccionar 9 jugadores para la alineación inicial?
- **b.** ¿Cuántas formas existen de seleccionar 9 jugadores para la alineación inicial y un orden al bat de los nueve inicialistas?
- **c.** Suponga que 5 de los 15 jugadores son zurdos. ¿Cuántas formas existen de seleccionar 3 jardineros zurdos y tener las otras 6 posiciones ocupadas por jugadores derechos?

34. Fallas del teclado de computadora pueden ser atribuidas a defectos eléctricos o mecánicos. Un taller de reparación actualmente cuenta con 25 teclados averiados, de los cuales 6 tienen defectos eléctricos y 19 tienen defectos mecánicos.
- **a.** ¿Cuántas maneras hay de seleccionar al azar cinco de estos teclados para una inspección completa (sin tener en cuenta el orden)?
- **b.** ¿De cuántas maneras puede seleccionarse una muestra de 5 teclados de manera que sólo dos tengan un defecto eléctrico?
- **c.** Si una muestra de 5 teclados se selecciona al azar, ¿cuál es la probabilidad de que al menos 4 de éstos tengan un defecto mecánico?

35. Una empresa de producción emplea 20 trabajadores en el turno de día, 15 en el turno de tarde y 10 en el turno de medianoche. Un consultor de control de calidad va a seleccionar 6 de estos trabajadores para entrevistas a fondo. Suponga que la selección se hace de tal modo que cualquier grupo particular de 6 trabajadores tiene la misma oportunidad de ser seleccionado al igual que cualquier otro grupo (sacando 6 papelitos de entre 45 sin reemplazarlos).
- **a.** ¿Cuántas selecciones resultarán en que los 6 trabajadores seleccionados provengan del turno de día? ¿Cuál es la probabilidad de que los 6 trabajadores seleccionados sean del turno de día?
- **b.** ¿Cuál es la probabilidad de que los 6 trabajadores seleccionados sean del mismo turno?
- **c.** ¿Cuál es la probabilidad de que por lo menos dos turnos diferentes estén representados entre los trabajadores seleccionados?

d. ¿Cuál es la probabilidad de que por lo menos uno de los turnos no esté representado en la muestra de trabajadores?

36. Un departamento académico compuesto de cinco profesores limitó su opción para jefe de departamento al candidato *A* o el candidato *B*. Cada miembro votó entonces con un papelito por uno de los candidatos. Suponga que en realidad existen tres votos para *A* y dos para *B*. Si los papelitos se cuentan al azar, ¿cuál es la probabilidad de que *A* permanezca adelante de *B* durante todo el conteo de votos? (P. ej. ¿ocurre este evento si el orden seleccionado es *AABAB* pero no si es *ABBAA*)?

37. Un experimentador está estudiando los efectos de la temperatura, la presión y el tipo de catalizador en la producción de cierta reacción química. Tres diferentes temperaturas, cuatro presiones distintas y cinco catalizadores diferentes se están considerando.

a. Si cualquier experimento particular implica utilizar una temperatura, una presión y un catalizador, ¿cuántos experimentos son posibles?

b. ¿Cuántos experimentos existen que impliquen el uso de la temperatura más baja y dos presiones bajas?

c. Suponga que se tienen que realizar cinco experimentos diferentes el primer día de experimentación. Si los cinco se eligen al azar de entre todas las posibilidades, de modo que cualquier grupo de cinco tenga la misma probabilidad de selección, ¿cuál es la probabilidad de que se utilice un catalizador diferente en cada experimento?

38. Una caja en un almacén contiene cuatro focos de 40 W, cinco de 60 W y seis de 75 W. Suponga que se eligen al azar tres focos.

a. ¿Cuál es la probabilidad de que exactamente dos de los focos seleccionados sean de 75 W?

b. ¿Cuál es la probabilidad de que los tres focos seleccionados sean de los mismos watts?

c. ¿Cuál es la probabilidad de que se seleccione un foco de cada tipo?

d. Suponga ahora que los focos tienen que ser seleccionados uno por uno hasta encontrar uno de 75 W. ¿Cuál es la probabilidad de que sea necesario examinar por lo menos seis focos?

39. Quince teléfonos acaban de llegar a un centro de servicio autorizado. Cinco de éstos son celulares, cinco inalámbricos y los otros cinco alámbricos. Suponga que a estos componentes se les asignan al azar los números 1, 2, . . . , 15, para establecer el orden en que serán reparados.

a. ¿Cuál es la probabilidad de que los teléfonos inalámbricos estén entre los primeros diez que van a ser reparados?

b. ¿Cuál es la probabilidad de que después de reparar diez de estos teléfonos, los teléfonos de sólo dos de los tres tipos queden para ser reparados?

c. ¿Cuál es la probabilidad de que dos teléfonos de cada tipo estén entre los primeros seis reparados?

40. Tres moléculas de tipo *A*, tres de tipo *B*, tres de tipo *C* y tres de tipo *D* tienen que ser unidas para formar una cadena molecular. Una cadena molecular como ésa es *ABCDABCDABCD* y otra es *BCDDAAABDBCC*.

a. ¿Cuántas moléculas de cadena hay? [*Sugerencia:* si las tres *A* se distinguen una de otra (A_1, A_2, A_3) y también las *B*, las *C* y las *D* ¿cuántas moléculas habría? ¿Cómo se reduce este número si se quitan los subíndices a las *A*?]

b. Supongamos que una molécula de cadena del tipo descrito se selecciona al azar. ¿Cuál es la probabilidad de que las tres moléculas de cada tipo terminen una al lado de la otra (como en *BBBAAADDDCCC*)?

41. Un número de identificación personal de cajeros automáticos (NIP) consta de cuatro cifras, cada una de 0, 1, 2, . . . 8, o 9, en sucesión.

a. ¿Cuántos números NIP posibles diferentes hay si no existen restricciones en la elección de dígitos?

b. De acuerdo con un representante en la sucursal local del autor del Chase Bank, hay restricciones en el hecho de la elección de dígitos. La opción es que se prohíba lo siguiente: (i) los cuatro dígitos idénticos (ii) las secuencias consecutivas de forma ascendente o descendente de dígitos, como 6543 (iii) cualquier secuencia de arranque con 19 (años de nacimiento son demasiado fáciles de adivinar). Así que si uno de los NIP en (a) es seleccionado al azar, ¿cuál es la probabilidad de que sea un NIP legítimo (es decir, no ser una de las secuencias prohibidas)?

c. Alguien ha robado una tarjeta de cajero automático y sabe que los dígitos primero y último del NIP son 8 y 1, respectivamente. Tiene tres intentos antes de que la tarjeta sea retenida por el cajero automático (pero no se da cuenta de eso). Así que selecciona al azar los dígitos 2º y 3º para el primer intento, a continuación, selecciona al azar un par de dígitos diferentes para el segundo intento, y otro par de dígitos seleccionados al azar para el tercer intento (el individuo sabe acerca de las restricciones descritas en (b) para seleccionar sólo de las posibilidades legítimas). ¿Cuál es la probabilidad de que el individuo tenga acceso a la cuenta?

d. Vuelva a calcular la probabilidad de (c) si los dígitos primero y último son 1 y 1, respectivamente.

42. Una alineación titular en el baloncesto se compone de dos defensas, dos delanteros y un centro.

a. Un equipo de la universidad tiene en su lista tres centros, cuatro defensas, cuatro delanteros y un individuo (X) que puede jugar tanto de defensa o como de delantero. ¿Cuántas alineaciones diferentes de inicio se pueden crear? [*Sugerencia*: considere la posibilidad de alineaciones sin X, luego alineaciones con X como defensa, a continuación, alineaciones con X como delantero.]

b. Ahora supongamos que la lista tiene 5 defensas, 5 delanteros, 3 centros y 2 "jugadores comodín" (X y Y), que pueden jugar tanto de defensas como delanteros. Si 5 de los 15 jugadores son seleccionados al azar, ¿cuál es la probabilidad de que constituyan una alineación de inicio legítima?

43. En un juego de póker de cinco cartas, una corrida se compone de cinco cartas con denominaciones adyacentes (p. ej. 9 de tréboles, 10 de corazones, joto de corazones, reina de espadas y rey de tréboles). Suponiendo que los ases pueden estar arriba o abajo, si le reparten una mano de cinco cartas, ¿cuál es la probabilidad de que sea una corrida con un 10 como carta alta? ¿Cuál es la probabilidad de que sea una corrida del mismo palo?

44. Demuestre que $\binom{n}{k} = \binom{n}{n-k}$. Dé una interpretación que implique subconjuntos.

2.4 Probabilidad condicional

Las probabilidades asignadas a varios eventos dependen de lo que se sabe sobre la situación experimental cuando se hace la asignación. Subsiguiente a la asignación inicial puede llegar a estar disponible información parcial pertinente al resultado del experimento. Tal información puede hacer que se revisen algunas de las asignaciones de probabilidad. Para un evento particular A, se ha utilizado $P(A)$ para representar la probabilidad asignada a A; ahora se considera $P(A)$ como la probabilidad original o no condicional del evento A.

En esta sección, se examina cómo afecta la información de que "un evento B ha ocurrido" a la probabilidad asignada a A. Por ejemplo, A podría referirse a un individuo que sufre una enfermedad particular en presencia de ciertos síntomas. Si se realiza un examen de sangre en el individuo y el resultado es negativo (B = examen de sangre negativo), entonces la probabilidad de que tenga la enfermedad cambiará (deberá reducirse, pero no a cero, puesto que los exámenes de sangre no son infalibles). Se utilizará la notación $P(A|B)$ para representar la **probabilidad condicional de A dado que el evento B haya ocurrido.** B es el "evento condicionante".

Por ejemplo, considérese el evento A en que un estudiante seleccionado al azar en su universidad obtuvo todas las clases deseadas durante el ciclo de inscripciones del semestre anterior. Presumiblemente $P(A)$ no es muy grande. Sin embargo, supóngase que el estudiante seleccionado es un atleta con prioridad de inscripción especial (el evento B). Entonces $P(A|B)$ deberá ser sustancialmente más grande que $P(A)$, aunque quizás aún no cerca de 1.

Ejemplo 2.24 En una planta se ensamblan componentes complejos en dos líneas de ensamble diferentes, A y A'. La línea A utiliza equipo más viejo que A', por lo que es un poco más lenta y menos confiable. Suponga que en un día dado la línea A ensambla 8 componentes, de los cuales 2 han sido identificados como defectuosos (B) y 6 como no defectuosos (B'), mientras que A' ha producido 1 componente defectuoso y 9 no defectuosos. Esta información se resume en la tabla adjunta.

		Condición	
		B	***B′***
Línea	***A***	2	6
	A′	1	9

Ajeno a esta información, el gerente de ventas selecciona al azar 1 de estos 18 componentes para una demostración. Antes de la demostración

$$P(\text{componente de la línea } A \text{ seleccionado}) = P(A) = \frac{N(A)}{N} = \frac{8}{18} = .44$$

No obstante, si el componente seleccionado resulta defectuoso, entonces el evento B ha ocurrido, por lo que el componente debe haber sido 1 de los 3 de la columna B de la tabla. Como estos 3 componentes son igualmente probables entre ellos mismos una vez que B ha ocurrido,

$$P(A|B) = \frac{2}{3} = \frac{2/18}{3/18} = \frac{P(A \cap B)}{P(B)} \tag{2.2}$$

■

En la ecuación (2.2), la probabilidad condicional está expresada como una razón de probabilidades incondicionales. El numerador es la probabilidad de la intersección de los dos eventos, en tanto que el denominador es la probabilidad del evento condicionante B. Un diagrama de Venn ilumina esta relación (figura 2.8).

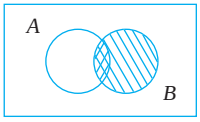


Figura 2.8 Motivación de la definición de probabilidad condicional

Dado que B ha ocurrido, el espacio muestral pertinente ya no es S pero consta de resultados en B; A ha ocurrido si y sólo si uno de los resultados en la intersección ocurrió, así que la probabilidad condicional de A dado B es proporcional a $P(A \cap B)$. Se utiliza la constante de proporcionalidad $1/P(B)$ para garantizar que la probabilidad $P(B|B)$ del nuevo espacio muestral B sea igual a 1.

Definición de probabilidad condicional

El ejemplo 2.24 demuestra que cuando los resultados son igualmente probables, el cálculo de probabilidades condicionales puede basarse en la intuición. Cuando los experimentos son más complicados, la intuición puede fallar, así que se requiere una definición general de probabilidad condicional que dé respuestas intuitivas en problemas simples. El diagrama de Venn y la ecuación (2.2) sugieren cómo proceder.

DEFINICIÓN

Para dos eventos cualesquiera A y B con $P(B) > 0$, la **probabilidad condicional de A dado que B ha ocurrido** está definida por

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{2.3}$$

Ejemplo 2.25 Supóngase que de todos los individuos que compran cierta cámara digital, 60% incluyen una tarjeta de memoria opcional en su compra, 40% incluyen una batería extra y 30% incluyen tanto una tarjeta como una batería. Considere seleccionar al azar un comprador y sean $A = \{\text{tarjeta de memoria adquirida}\}$ y $B = \{\text{batería adquirida}\}$. Entonces $P(A) = .60$, $P(B) = .40$ y $P(\text{ambas adquiridas}) = P(A \cap B) = .30$. Dado que el individuo seleccionado adquirió una batería extra, la probabilidad de que una tarjeta opcional también sea adquirida es

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{.30}{.40} = .75$$

Es decir, de todos los que adquieren una batería extra, 75% adquirieron una tarjeta de memoria opcional. Asimismo,

$$P(\text{batería}|\text{tarjeta de memoria}) = P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{.30}{.60} = .50$$

Obsérvese que $P(A|B) \neq P(A)$ y $P(B|A) \neq P(B)$. ■

El evento cuya probabilidad es deseada podría ser una unión o intersección de otros eventos y lo mismo podría ser cierto del evento condicionante.

Ejemplo 2.26 Una revista de noticias publica tres columnas tituladas "Arte" (A), "Libros" (B) y "Cine" (C). Los hábitos de lectura de un lector seleccionado al azar con respecto a estas columnas son

Lee regularmente	A	B	C	$A \cap B$	$A \cap C$	$B \cap C$	$A \cap B \cap C$
Probabilidad	.14	.23	.37	.08	.09	.13	.05

La figura 2.9 ilustra las probabilidades pertinentes.

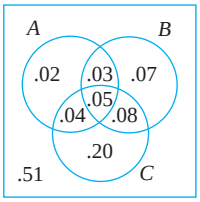


Figura 2.9 Diagrama de Venn para el ejemplo 2.26

Por lo tanto se tiene

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{.08}{.23} = .348$$

$$P(A|B \cup C) = \frac{P(A \cap (B \cup C))}{P(B \cup C)} = \frac{.04 + .05 + .03}{.47} = \frac{.12}{.47} = .255$$

$$P(A|\text{lee por lo menos una}) = P(A|A \cup B \cup C) = \frac{P(A \cap (A \cup B \cup C))}{P(A \cup B \cup C)}$$

$$= \frac{P(A)}{P(A \cup B \cup C)} = \frac{.14}{.49} = .286$$

y

$$P(A \cup B|C) = \frac{P((A \cup B) \cap C)}{P(C)} = \frac{.04 + .05 + .08}{.37} = .459$$

■

Regla de multiplicación para $P(A \cap B)$

La definición de probabilidad condicional da el siguiente resultado, obtenido multiplicando ambos miembros de la ecuación (2.3) por $P(B)$.

La regla de multiplicación

$$P(A \cap B) = P(A|B) \cdot P(B)$$

Esta regla es importante porque a menudo es el caso de que se desea $P(A \cap B)$, en tanto que $P(B)$ y $P(A|B)$ pueden ser especificadas a partir de la descripción del problema. La consideración de $P(B|A)$ da $P(A \cap B) = P(B|A) \cdot P(A)$.

Ejemplo 2.27 Cuatro individuos han respondido a una solicitud de un banco de sangre para efectuar donaciones. Ninguno de ellos ha donado antes, por lo que sus tipos de sangre son desconocidos. Suponga que sólo se desea el tipo O+ y sólo uno de los cuatro tiene ese tipo. Si los donadores potenciales se seleccionan en orden aleatorio para determinar su tipo de sangre, ¿cuál es la probabilidad de que por los menos tres individuos tengan que ser examinados para determinar su tipo de sangre para obtener el tipo deseado?

Haciendo la identificación $B = \{\text{primer tipo no O+}\}$ y $A = \{\text{segundo tipo no O+}\}$, $P(B) = \frac{3}{4}$. Dado que el primer tipo no es O+, dos de los tres individuos que quedan no son O+, por lo tanto $P(A|B) = \frac{2}{3}$. La regla de multiplicación ahora da

$$\begin{aligned} P(\text{por lo menos tres individuos fueron examinados para determinar su tipo de sangre}) &= P(A \cap B) \\ &= P(A|B) \cdot P(B) \\ &= \frac{2}{3} \cdot \frac{3}{4} = \frac{6}{12} \\ &= .5 \end{aligned}$$

■

La regla de multiplicación es más útil cuando los experimentos se componen de varias etapas en sucesión. El evento condicionante B describe entonces el resultado de la primera etapa y A el resultado de la segunda, de modo que $P(A|B)$, condicionada en lo que ocurra primero, a menudo será conocida. La regla es fácil de ser ampliada a experimentos que implican más de dos etapas. Por ejemplo,

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_3|A_1 \cap A_2) \cdot P(A_1 \cap A_2) \\ &= P(A_3|A_1 \cap A_2) \cdot P(A_2|A_1) \cdot P(A_1) \end{aligned} \qquad (2.4)$$

donde A_1 ocurre primero, seguido por A_2 y finalmente A_3.

Ejemplo 2.28 Para el experimento de determinación de tipo de sangre del ejemplo 2.27,

$$\begin{aligned} P(\text{el tercer tipo es O+}) &= P(\text{el tercero es} \mid \text{el primero no es} \cap \text{el segundo no es}) \\ &\quad \cdot P(\text{el segundo no es} \mid \text{el primero no es}) \cdot P(\text{el primero no es}) \\ &= \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{4} = \frac{1}{4} = .25 \end{aligned}$$

■

Cuando el experimento de interés se compone de una secuencia de varias etapas, es conveniente representarlas con un diagrama de árbol. Una vez que se tiene un diagrama de árbol apropiado, las probabilidades y las probabilidades condicionales pueden ser ingresadas en las diversas ramas; esto implicará el uso repetido de la regla de multiplicación.

Ejemplo 2.29 Una cadena de tiendas de video vende tres marcas diferentes de reproductores de DVD. De sus ventas de reproductores de DVD, 50% son de la marca 1 (la menos cara), 30% son de la marca 2 y 20% son de la marca 3. Cada fabricante ofrece 1 año de garantía en las partes y mano de obra. Se sabe que 25% de los reproductores de DVD de la marca 1 requieren trabajo de reparación dentro del periodo de garantía, mientras que los porcentajes correspondientes de las marcas 2 y 3 son 20% y 10%, respectivamente.

1. ¿Cuál es la probabilidad de que un comprador seleccionado al azar haya adquirido un reproductor de DVD marca 1 que necesitará reparación mientras se encuentra dentro de la garantía?
2. ¿Cuál es la probabilidad de que un comprador seleccionado al azar haya comprado un reproductor de DVD que necesitará reparación mientras se encuentra dentro de la garantía?
3. Si un cliente regresa a la tienda con un reproductor de DVD que necesita reparación dentro de la garantía, ¿cuál es la probabilidad de que sea un reproductor de DVD marca 1? ¿Un reproductor de DVD marca 2? ¿Un reproductor de DVD marca 3?

La primera etapa del problema implica un cliente que selecciona una de las tres marcas de reproductor de DVD. Sea A_i = {marca i adquirida}, con i = 1, 2, y 3. Entonces $P(A_1) = .50$, $P(A_2) = .30$ y $P(A_3) = .20$. Una vez que se selecciona una marca de reproductor de DVD, la segunda etapa implica observar si el reproductor de DVD seleccionado necesita reparación dentro de la garantía. Con B = {necesita reparación} y B' = {no necesita reparación}, la información dada implica que $P(B|A_1) = .25$, $P(B|A_2) = .20$ y $P(B|A_3) = .10$.

El diagrama de árbol que representa esta situación experimental se muestra en la figura 2.10. Las ramas iniciales corresponden a marcas diferentes de reproductores de DVD; hay dos ramas de segunda generación que emanan de la punta de cada rama inicial, una para "necesita reparación" y la otra para "no necesita reparación". La probabilidad $P(A_i)$ aparece en la rama i-ésima inicial, en tanto que las probabilidades condicionales $P(B|A_i)$ y $P(B'|A_i)$ aparecen en las ramas de la segunda generación. A la derecha de cada rama de segunda generación correspondiente a la ocurrencia de B, se muestra el producto de probabilidades en las ramas que conducen hacia fuera de dicho punto. Ésta es simplemente la regla de multiplicación en acción. La respuesta a la pregunta planteada en 1 es por lo tanto $P(A_1 \cap B) = P(B|A_1) \cdot P(A_1) = .125$. La respuesta a la pregunta 2 es

$$\begin{aligned} P(B) &= P[\text{(marca 1 y reparación) o (marca 2 y reparación) o (marca 3 y reparación)}] \\ &= P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B) \\ &= .125 + .060 + .020 = .205 \end{aligned}$$

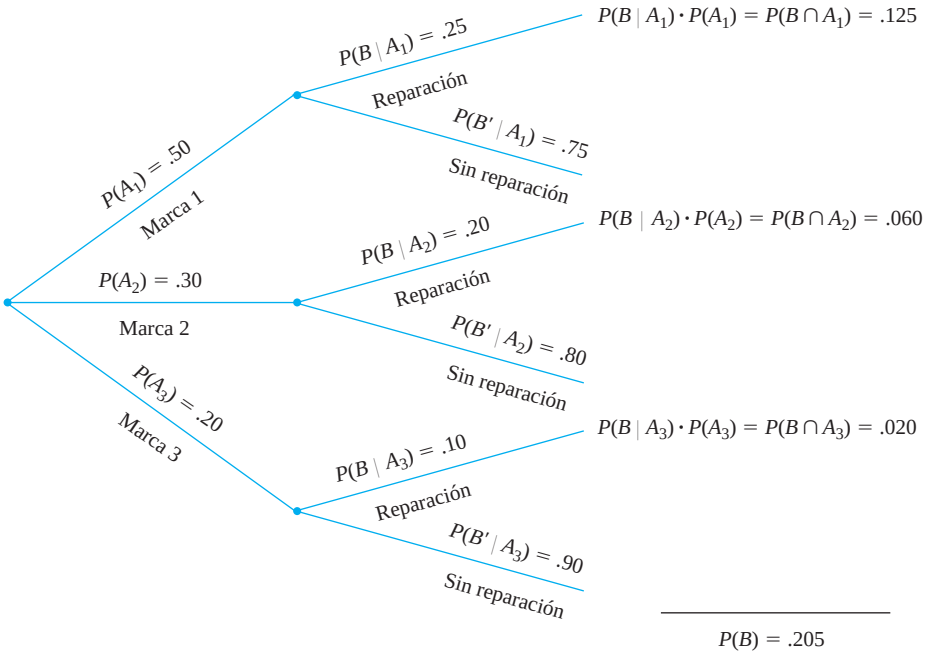


Figura 2.10 Diagrama de árbol para el ejemplo 2.29

Finalmente,

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{.125}{.205} = .61$$

$$P(A_2|B) = \frac{P(A_2 \cap B)}{P(B)} = \frac{.060}{.205} = .29$$

y

$$P(A_3|B) = 1 - P(A_1|B) - P(A_2|B) = .10$$

La *probabilidad previa* o inicial de la marca 1 es .50. Una vez que se sabe que el reproductor de DVD seleccionado necesitaba reparación, la *probabilidad posterior* de la marca 1 se incrementa a .61. Esto se debe a que es más probable que los reproductores de DVD marca 1 necesiten reparación de garantía que las demás marcas. La probabilidad posterior de la marca 3 es $P(A_3 \mid B) = .10$, la cual es mucho menor que la probabilidad previa $P(A_3) = .20$. ■

Teorema de Bayes

El cálculo de una probabilidad posterior $P(A_j|B)$ a partir de probabilidades previas dadas $P(A_i)$ y probabilidades condicionales $P(B|A_i)$ ocupa una posición central en la probabilidad elemental. La regla general de dichos cálculos, los que en realidad son una aplicación simple de la regla de multiplicación, se remonta al reverendo Thomas Bayes, quien vivió en el siglo XVIII. Para formularla primero se requiere otro resultado. Recuérdese que los eventos $A_1, \ldots, A_k$ son mutuamente exclusivos si ninguno de los dos tiene resultados comunes. Los eventos son *exhaustivos* si un A_i debe ocurrir, de modo que $A_1 \cup \ldots \cup A_k = \mathcal{S}$.

Ley de probabilidad total

Sean $A_1, \ldots, A_k$ eventos mutuamente exclusivos y exhaustivos. Entonces para cualquier otro evento B,

$$\begin{aligned} P(B) &= P(B|A_1)P(A_1) + \cdots + P(B|A_k)P(A_k) \\ &= \sum_{i=1}^{k} P(B|A_i)P(A_i) \end{aligned} \tag{2.5}$$

Comprobación Como los eventos A_i son mutuamente exclusivos y exhaustivos, si B ocurre debe ser en forma conjunta con uno de los eventos A_i de manera exacta. Es decir, $B = (A_1 \cap B) \cup \ldots \cup (A_k \cap B)$, donde los eventos $(A_i \cap B)$ son mutuamente excluyentes. Esta "partición de B" se ilustra en la figura 2.11. Por lo tanto

$$P(B) = \sum_{i=1}^{k} P(A_i \cap B) = \sum_{i=1}^{k} P(B|A_i)P(A_i)$$

como se deseaba.

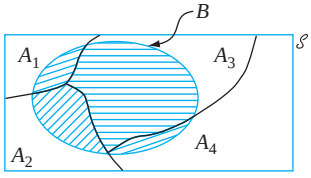


Figura 2.11 Partición de B entre las A_i' mutuamente excluyentes y exhaustivas. ■

Ejemplo 2.30 Una persona tiene 3 cuentas de correo electrónico diferentes. La mayoría de sus mensajes, el 70%, entra en la cuenta # 1, mientras que el 20% entra en la cuenta # 2 y el 10% restante en la cuenta # 3. De los mensajes en la cuenta # 1, sólo el 1% son spam, mientras que los porcentajes correspondientes a las cuentas # 2 y # 3 son 2% y 5%, respectivamente. ¿Cuál es la probabilidad de que un mensaje spam sea seleccionado al azar?

Para responder a esta pregunta, primero vamos a establecer una notación:

$A_i = \{\text{el mensaje es de la cuenta} \neq i\}$ para $i = 1, 2, 3$, $B = \{\text{el mensaje es spam}\}$

Entonces, los porcentajes dados implican que

$$P(A_1) = .70,\ P(A_2) = .20,\ P(A_3) = .10$$
$$P(B|A_1) = .01,\ P(B|A_2) = .02,\ P(B|A_3) = .05$$

Ahora bien, es simplemente una cuestión de sustituir en la ecuación de la ley de probabilidad total:

$$P(B) = (.01)(.70) + (.02)(.20) + (.05)(.10) = .016$$

A largo plazo, el 1.6% de los mensajes de esta persona serán spam. ■

Teorema de Bayes

Sean $A_1, A_2, \ldots, A_k$ un conjunto de eventos mutuamente exclusivos y exhaustivos con probabilidades *previas* $P(A_i)$ $(i = 1, \ldots, k)$. Entonces para cualquier otro evento B para el cual $P(B) > 0$, la probabilidad *posterior* de A_j dado que B ha ocurrido es

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{k} P(B|A_i) \cdot P(A_i)} \quad j = 1, \ldots, k \qquad (2.6)$$

La transición de la segunda a la tercera expresión en (2.6) se apoya en el uso de la regla de multiplicación en el numerador y la ley de probabilidad total en el denominador. La proliferación de eventos y subíndices en (2.6) puede ser un poco intimidante para los recién llegados a la probabilidad. Mientras existan relativamente pocos eventos en la repartición, se puede utilizar un diagrama de árbol (como en el ejemplo 2.29) como base para calcular probabilidades posteriores sin jamás referirse de manera explícita al teorema de Bayes.

Ejemplo 2.31 *Incidencia de una enfermedad rara*. Sólo 1 de 1000 adultos padece una enfermedad rara para la cual se ha creado una prueba de diagnóstico. La prueba es tal que cuando un individuo en realidad tiene la enfermedad, un resultado positivo se presentará en 99% de las veces mientras que en individuos sin enfermedad el examen será positivo sólo el 2% de las veces. Si se somete a prueba un individuo seleccionado al azar y el resultado es positivo, ¿cuál es la probabilidad de que el individuo tenga la enfermedad?

Para utilizar el teorema de Bayes, sea A_1 = el individuo tiene la enfermedad, A_2 = el individuo no tiene la enfermedad y B = resultado de prueba positivo. Entonces $P(A_1) = .001$, $P(A_2) = .999$, $P(B|A_1) = .99$ y $P(B|A_2) = .02$. El diagrama de árbol para este problema aparece en la figura 2.12.

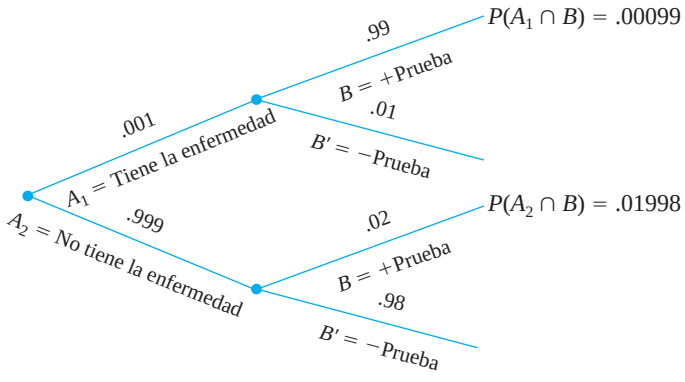


Figure 2.12 Diagrama de árbol para el problema de una enfermedad rara

Junto a cada rama correspondiente a un resultado positivo de prueba, la regla de multiplicación da las probabilidades anotadas. Por consiguiente, $P(B) = .00099 + .01998 = .02097$, a partir de la cual se tiene

$$P(A_1 \mid B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{.00099}{.02097} = .047$$

Este resultado parece contraintuitivo; la prueba de diagnóstico parece tan precisa que es altamente probable que alguien con un resultado positivo de prueba tenga la enfermedad, mientras que la probabilidad condicional calculada es de sólo .047. Sin embargo, como la enfermedad es rara y la prueba es sólo moderadamente confiable, surgen más resultados positivos de prueba a causa de errores y no de individuos enfermos. La probabilidad de tener la enfermedad se ha incrementado por un factor de multiplicación de 47 (desde la probabilidad previa de .001 hasta la probabilidad posterior de .047); pero para incrementar aún más la probabilidad posterior, se requiere una prueba de diagnóstico con tasas de error mucho más pequeñas. ■

EJERCICIOS Sección 2.4 (45–69)

45. La población de un país particular se compone de tres grupos étnicos. Cada individuo pertenece a uno de los cuatro grupos sanguíneos principales. La *tabla de probabilidad conjunta* anexa da la proporción de individuos en las diversas combinaciones de grupo étnico–grupo sanguíneo.

		Grupo sanguíneo			
		O	**A**	**B**	**AB**
Grupo étnico	**1**	.082	.106	.008	.004
	2	.135	.141	.018	.006
	3	.215	.200	.065	.020

Suponga que se selecciona un individuo al azar de la población y que los eventos se definen como $A = \{\text{tipo A seleccionado}\}$, $B = \{\text{tipo B seleccionado}\}$ y $C = \{\text{grupo étnico 3 seleccionado}\}$.

a. Calcule $P(A)$, $P(C)$ y $P(A \cap C)$.

b. Calcule tanto $P(A \mid C)$ y $P(C \mid A)$ y explique en contexto lo que cada una de estas probabilidades representa.

c. Si el individuo seleccionado no tiene sangre de tipo B, ¿cuál es la probabilidad de que él o ella pertenezcan al grupo étnico 1?

46. Suponga que un individuo es seleccionado al azar de la población de todos los adultos varones que viven en Estados Unidos. Sea A el evento en que el individuo seleccionado tiene una estatura de más de 6 pies, y sea B el evento en que el individuo seleccionado es un jugador profesional de basquetbol. ¿Cuál piensa que es más grande, $P(A \mid B)$ o $P(B \mid A)$? ¿Por qué?

47. Regrese al escenario de la tarjeta de crédito del ejercicio 12 (sección 2.2), donde $A = \{\text{Visa}\}$, $B = \{\text{MasterCard}\}$, $P(A) = .5$, $P(B) = .4$ y $P(A \cap B) = .25$. Calcule e interprete cada una de las siguientes probabilidades (un diagrama de Venn podría ayudar).

a. $P(B \mid A)$ **b.** $P(B' \mid A)$

c. $P(A \mid B)$ **d.** $P(A' \mid B)$

e. Dado que el individuo seleccionado tiene por lo menos una tarjeta, ¿cuál es la probabilidad de que él o ella tengan una tarjeta Visa?

48. Reconsidere la situación del sistema defectuoso descrito en el ejercicio 26 (sección 2.2).

a. Dado que el sistema tiene un defecto de tipo 1, ¿cuál es la probabilidad de que tenga un defecto de tipo 2?

b. Dado que el sistema tiene un defecto de tipo 1, ¿cuál es la probabilidad de que tenga los tres tipos de defecto?

c. Dado que el sistema tiene por lo menos un tipo de defecto, ¿cuál es la probabilidad de que tenga exactamente un tipo de defecto?

d. Dado que el sistema tiene los primeros dos tipos de defecto, ¿cuál es la probabilidad de que no tenga el tercer tipo de defecto?

49. La tabla adjunta proporciona información sobre el tipo de café seleccionado por alguien que compra una taza en un kiosco del aeropuerto en particular

	Pequeño	**Mediano**	**Grande**
Regular	14%	20%	26%
Descafeinado	20%	10%	10%

Considere la posibilidad de seleccionar al azar un comprador de café.

a. ¿Cuál es la probabilidad de que la persona adquiera una taza pequeña? ¿Una taza de café descafeinado?

b. Si nos enteramos de que la persona seleccionada compra una taza pequeña, ¿cuál es ahora la probabilidad de que él/ella escoja el café descafeinado y cómo interpreta esta probabilidad?

c. Si nos enteramos de que el individuo seleccionado compró un café descafeinado, ¿cuál es ahora la probabilidad de que un tamaño pequeño fue el escogido, y cómo se compara esto con la probabilidad incondicional correspondiente de (a)?

50. Una tienda de departamentos vende camisas sport en tres tallas (pequeña, mediana y grande), tres diseños (a cuadros, estampadas y a rayas) y dos largos de manga (larga y corta). Las tablas adjuntas dan las proporciones de camisas vendidas en las varias combinaciones de categoría.

Manga corta

	Diseño		
Talla	**A cuadros**	**Estampada**	**Rayas**
Ch	.04	.02	.05
M	.08	.07	.12
G	.03	.07	.08

Manga larga

	Diseño		
Talla	**A cuadros**	**Estampada**	**Rayas**
Ch	.03	.02	.03
M	.10	.05	.07
G	.04	.02	.08

a. ¿Cuál es la probabilidad de que la siguiente camisa vendida sea una camisa mediana, estampada, de manga larga?
b. ¿Cuál es la probabilidad de que la siguiente camisa vendida sea una camisa estampada, mediana?
c. ¿Cuál es la probabilidad de que la siguiente camisa vendida sea de manga corta? ¿De manga larga?
d. ¿Cuál es la probabilidad de que la talla de la siguiente camisa vendida sea mediana? ¿Que la siguiente camisa vendida sea estampada?
e. Dado que la camisa que se acaba de vender era de manga corta a cuadros, ¿cuál es la probabilidad de que fuera mediana?
f. Dado que la camisa que se acaba de vender era mediana a cuadros, ¿cuál es la probabilidad de que fuera de manga corta? ¿De manga larga?

51. Una caja contiene seis pelotas rojas y cuatro verdes y una segunda caja contiene siete pelotas rojas y tres verdes. Se selecciona una pelota al azar de la primera caja y se la coloca en la segunda caja. Luego se selecciona al azar una pelota de la segunda caja y se la coloca en la primera caja.
a. ¿Cuál es la probabilidad de que se seleccione una pelota roja de la primera caja y de que se seleccione una pelota roja de la segunda caja?
b. Al final del proceso de selección, ¿cuál es la probabilidad de que los números de pelotas rojas y verdes que hay en la primera caja sean idénticos a los números iniciales?

52. Un sistema se compone de bombas idénticas, #1 y #2. Si una falla, el sistema seguirá operando. Sin embargo, debido al esfuerzo adicional, ahora es más probable que la bomba restante falle de lo que era originalmente. Es decir, $r = P(\#2 \text{ falla} \mid \#1 \text{ falla}) > P(\#2 \text{ falla}) = q$. Si por lo menos una bomba falla alrededor del final de su vida útil en 7% de todos los sistemas y ambas bombas fallan durante dicho periodo en sólo 1%, ¿cuál es la probabilidad de que la bomba #1 falle durante su vida útil?

53. Un taller repara componentes tanto de audio como de video. Sea A el evento en que el siguiente componente traído a reparación es un componente de audio, y sea B el evento en que el siguiente componente es un reproductor de discos compactos (así que el evento B está contenido en A). Suponga que $P(A) = .6$ y $P(B) = .05$. ¿Cuál es $P(B \mid A)$?

54. En el ejercicio 13, $A_i = \{\text{proyecto otorgado } i\}$, con $i = 1, 2, 3$. Use las probabilidades dadas allí para calcular las siguientes probabilidades y explique en palabras el significado de cada una.
a. $P(A_2 \mid A_1)$ **b.** $P(A_2 \cap A_3 \mid A_1)$
c. $P(A_2 \cup A_3 \mid A_1)$ **d.** $P(A_1 \cap A_2 \cap A_3 \mid A_1 \cup A_2 \cup A_3)$.

55. Las garrapatas de venados pueden ser portadoras de la enfermedad de Lyme o de la erliquiosis granulocítica humana (HGE, por sus siglas en inglés). Con base en un estudio reciente, suponga que 16% de todas las garrapatas en cierto lugar portan la enfermedad de Lyme, 10% portan HGE y 10% de las garrapatas que portan por lo menos una de estas enfermedades en realidad portan las dos. Si se determina que una garrapata seleccionada al azar ha sido portadora de HGE, ¿cuál es la probabilidad de que la garrapata seleccionada también porte la enfermedad de Lyme?

56. Para los eventos A y B con $P(B) > 0$, demuestre que $P(A \mid B) + P(A' \mid B) = 1$.

57. Si $P(B \mid A) > P(B)$, demuestre que $P(B' \mid A) < P(B')$. [*Sugerencia*: sume $P(B' \mid A)$ a ambos lados de la desigualdad dada y luego utilice el resultado del ejercicio 56.]

58. Demuestre que para tres eventos cualesquiera A, B y C con $P(C) > 0$, $P(A \cup B \mid C) = P(A \mid C) + P(B \mid C) - P(A \cap B \mid C)$.

59. En una gasolinera, 40% de los clientes utilizan gasolina regular (A_1), 35% usan gasolina plus (A_2) y 25% utilizan premium (A_3). De los clientes que utilizan gasolina regular, sólo 30% llenan sus tanques (evento B). De los clientes que utilizan plus, 60% llenan sus tanques, mientras que los que utilizan premium, 50% llenan sus tanques.
a. ¿Cuál es la probabilidad de que el siguiente cliente pida gasolina plus y llene el tanque ($A_2 \cap B$)?
b. ¿Cuál es la probabilidad de que el siguiente cliente llene el tanque?
c. Si el siguiente cliente llena el tanque, ¿cuál es la probabilidad que pida gasolina regular? ¿Plus? ¿Premium?

60. 70% de las aeronaves ligeras que desaparecen en vuelo en cierto país son posteriormente localizadas. De las aeronaves que son localizadas, 60% cuentan con un localizador de emergencia, mientras que 90% de las aeronaves no localizadas no cuentan con dicho localizador. Suponga que una aeronave ligera ha desaparecido.
a. Si tiene un localizador de emergencia, ¿cuál es la probabilidad de que no será localizada?
b. Si no tiene un localizador de emergencia, ¿cuál es la probabilidad de que será localizada?

61. Componentes de cierto tipo son enviados a un distribuidor en lotes de diez. Suponga que 50% de dichos lotes no contienen componentes defectuosos, 30% contienen un componente defectuoso y 20% contienen dos componentes defectuosos. Se seleccionan al azar dos componentes de un lote y se prueban.

¿Cuáles son las probabilidades asociadas con 0, 1 y 2 componentes defectuosos que están en el lote en cada una de las siguientes condiciones?

a. Ningún componente probado está defectuoso.

b. Uno de los dos componentes probados está defectuoso. [*Sugerencia*: trace un diagrama de árbol con tres ramas de primera generación correspondientes a los tres tipos diferentes de lotes.]

62. Una compañía que fabrica cámaras de video produce un modelo básico y un modelo de lujo. Durante el año pasado, 40% de las cámaras vendidas fueron del modelo básico. De aquellos que compraron el modelo básico, 30% adquirieron una garantía ampliada, en tanto que 50% de los que compraron el modelo de lujo también lo hicieron. Si se sabe que un comprador seleccionado al azar tiene una garantía ampliada, ¿qué tan probable es que él o ella tengan un modelo básico?

63. Para los clientes que compran un refrigerador en una tienda de aparatos domésticos, sea A el evento en que el refrigerador fue fabricado en EU, B el evento en que el refrigerador contaba con una máquina de hacer hielos y C el evento en que el cliente adquirió una garantía ampliada. Las probabilidades pertinentes son

$$P(A) = .75 \quad P(B|A) = .9 \quad P(B|A') = .8$$
$$P(C|A \cap B) = .8 \quad P(C|A \cap B') = .6$$
$$P(C|A' \cap B) = .7 \quad P(C|A' \cap B') = .3$$

a. Construya un diagrama de árbol compuesto de ramas de primera, segunda y tercera generaciones y anote el evento y la probabilidad apropiada junto a cada rama.

b. Calcule $P(A \cap B \cap C)$.

c. Calcule $P(B \cap C)$.

d. Calcule $P(C)$.

e. Calcule $P(A|B \cap C)$, la probabilidad de la compra de un refrigerador fabricado en EU dado que también se adquirieron una máquina de hacer hielos y una garantía ampliada.

64. El editor de comentarios de una cierta revista científica decide si la revisión de cualquier libro en particular debe ser corta (1–2 páginas), mediana (3–4 páginas), o larga (5–6 páginas). Los datos sobre estudios recientes indican que el 60% de ellas son cortas, el 30% son medianas y el otro 10% son largas. Los comentarios están presentados en Word o LaTeX. Para las revisiones cortas, el 80% son en Word, mientras que el 50% de las revisiones medianas se encuentran en Word y el 30% de las revisiones largas están en Word. Supongamos que se selecciona aleatoriamente una revisión reciente.

a. ¿Cuál es la probabilidad de que la revisión seleccionada se presentara en formato Word?

b. Si la revisión seleccionada se presentó en formato Word, ¿cuáles son las probabilidades posteriores de que sea corto, mediano o largo?

65. Un gran operador de complejos de tiempo compartido requiere que cualquier persona interesada en hacer una compra primero visite el sitio de interés. Los datos históricos indican que el 20% de todos los compradores potenciales seleccionaron un día de visita, el 50% elige una visita de una noche y el 30% opta por una visita de dos noches. Además, el 10% de los visitantes de un día en última instancia, hacen una compra, el 30% de los visitantes de una noche compran una unidad y el 20% de los visitantes de dos noches deciden comprar. Supongamos que un visitante es seleccionado al azar y se demuestra que ha realizado una compra. ¿Qué tan probable es que esta persona haya realizado una visita de día? ¿Una visita de una noche? ¿Una visita de dos noches?

66. Considere la siguiente información sobre vacacionistas (basada en parte en una encuesta reciente de Travelocity): 40% revisan su correo electrónico de trabajo, 30% utilizan un teléfono celular para permanecer en contacto con su trabajo, 25% trajeron una computadora portátil consigo, 23% revisan su correo electrónico de trabajo y utilizan un teléfono celular para permanecer en contacto, y 51% ni revisan su correo electrónico de trabajo ni utilizan un teléfono celular para permanecer en contacto ni trajeron consigo una computadora portátil. Además, 88 de cada 100 que traen una computadora portátil también revisan su correo electrónico de trabajo y 70 de cada 100 que utilizan un teléfono celular para permanecer en contacto también traen una computadora portátil.

a. ¿Cuál es la probabilidad de que un vacacionista seleccionado al azar que revisa su correo electrónico de trabajo también utilice un teléfono celular para permanecer en contacto?

b. ¿Cuál es la probabilidad de que alguien que trae una computadora portátil también utilice un teléfono celular para permanecer en contacto?

c. Si el vacacionista seleccionado al azar revisó su correo electrónico de trabajo y trajo una computadora portátil, ¿cuál es la probabilidad de que él o ella utilice un teléfono celular para permanecer en contacto?

67. Ha habido una gran controversia durante los últimos años con respecto a qué tipos de vigilancia son apropiados para impedir el terrorismo. Suponga que un sistema de vigilancia particular tiene 99% de probabilidades de identificar correctamente a un futuro terrorista y 99.9% de probabilidades de identificar correctamente a alguien que no es un futuro terrorista. Si existen 1000 futuros terroristas en una población de 300 millones y se selecciona al azar uno de estos 300 millones, es examinado por el sistema e identificado como futuro terrorista, ¿cuál es la probabilidad de que él o ella sea en realidad un futuro terrorista? ¿Le inquieta el valor de esta probabilidad sobre el uso del sistema de vigilancia? Explique.

68. Una amiga que vive en Los Ángeles hace viajes frecuentes de consultoría a Washington, D.C.; 50% del tiempo viaja en la línea aérea #1, 30% del tiempo en la aerolínea #2 y el 20% restante en la aerolínea #3. Los vuelos de la aerolínea #1 llegan demorados a D.C. 30% del tiempo y 10% del tiempo llegan demorados a L.A. Para la aerolínea #2, estos porcentajes son 25% y 20%, en tanto que para la aerolínea #3 los porcentajes son 40% y 25%. Si se sabe que en un viaje particular ella llegó demorada a exactamente uno de los dos destinos, ¿cuáles son las probabilidades posteriores de haber volado en las aerolíneas #1, #2 y #3? Suponga que la probabilidad de arribar con demora a L.A. no se ve afectada por lo que suceda en el vuelo a D.C. [*Sugerencia*: desde la punta de cada rama de primera generación en un diagrama de árbol, trace tres ramas de segunda generación identificadas, respectivamente, como 0 demorado, 1 demorado y 2 demorado.]

69. En el ejercicio 59, considere la siguiente información adicional sobre el uso de tarjetas de crédito:

70% de todos los clientes que utilizan gasolina regular y que llenan el tanque usan una tarjeta de crédito.

50% de todos los clientes que utilizan gasolina regular y que no llenan el tanque usan una tarjeta de crédito.
60% de todos los clientes que llenan el tanque con gasolina plus usan una tarjeta de crédito.
50% de todos los clientes que utilizan gasolina plus y que no llenan el tanque usan una tarjeta de crédito.
50% de todos los clientes que utilizan gasolina premium y que llenan el tanque usan una tarjeta de crédito.
40% de todos los clientes que utilizan gasolina premium y que no llenan el tanque usan una tarjeta de crédito.

Calcule la probabilidad de cada uno de los siguientes eventos para el siguiente cliente que llegue (un diagrama de árbol podría ayudar).

a. {plus, tanque lleno y tarjeta de crédito}
b. {premium, tanque no lleno y tarjeta de crédito}
c. {premium y tarjeta de crédito}
d. {tanque lleno y tarjeta de crédito}
e. {tarjeta de crédito}
f. Si el siguiente cliente utiliza una tarjeta de crédito, ¿cuál es la probabilidad de que pida premium?

2.5 Independencia

La definición de probabilidad condicional permite revisar la probabilidad $P(A)$ originalmente asignada a A cuando después se informa que otro evento B ha ocurrido; la nueva probabilidad de A es $P(A|B)$. En los ejemplos, con frecuencia fue el caso de que $P(A|B)$ difería de la probabilidad no condicional $P(A)$, lo que indica que la información "B ha ocurrido" cambia la probabilidad de que ocurra A. A menudo la probabilidad de que ocurra o haya ocurrido A no se ve afectada por el conocimiento de que B ha ocurrido, así que $P(A|B) = P(A)$. Es entonces natural considerar a A y B como eventos independientes, es decir que la ocurrencia o no ocurrencia de un evento no afecta la probabilidad de que el otro ocurra.

DEFINICIÓN

Los eventos A y B son **independientes** si $P(A|B) = P(A)$ y son **dependientes** en caso contrario.

La definición de independencia podría parecer "no simétrica" porque no demanda también que $P(B|A) = P(B)$. Sin embargo, utilizando la definición de probabilidad condicional y la regla de multiplicación,

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B)P(B)}{P(A)} \tag{2.7}$$

El lado derecho de la ecuación (2.7) es $P(B)$ si y sólo si $P(A|B) = P(A)$ (independencia), así que la igualdad en la definición implica la otra igualdad (y viceversa). También es fácil demostrar que si A y B son independientes, entonces también lo son los pares de eventos: (1) A' y B, (2) A y B' y (3) A' y B'.

Ejemplo 2.32 Considere una gasolinera con seis bombas numeradas 1, 2, . . . , 6, y sea E_i el evento simple en que un cliente seleccionado al azar utiliza la bomba i ($i = 1, \ldots, 6$). Suponga que

$$P(E_1) = P(E_6) = .10, \quad P(E_2) = P(E_5) = .15, \quad P(E_3) = P(E_4) = .25$$

Defina los eventos A, B, C como

$$A = \{2, 4, 6\}, B = \{1, 2, 3\}, C = \{2, 3, 4, 5\}.$$

Luego se tiene $P(A) = .50$, $P(A|B) = .30$ y $P(A|C) = .50$. Es decir, los eventos A y B son dependientes, en tanto que los eventos A y C son independientes. Intuitivamente, A y C son independientes porque la división de probabilidad relativa entre las bombas pares e impares es la misma entre las bombas 2, 3, 4, 5, como lo es entre todas las seis bombas. ■

Ejemplo 2.33 Sean A y B dos eventos mutuamente exclusivos cualesquiera con $P(A) > 0$. Por ejemplo, para un automóvil seleccionado al azar, sea $A = \{$el carro es de cuatro cilindros$\}$ y $B = \{$el carro es de seis cilindros$\}$. Como los eventos son mutuamente exclusivos, si B ocurre, entonces A quizás puede no haber ocurrido, así que $P(A|B) = 0 \neq P(A)$. El mensaje aquí es que *si dos eventos son mutuamente exclusivos, no pueden ser independientes*.

Cuando A y B son mutuamente exclusivos, la información de que A ocurrió dice algo sobre B (no puede haber ocurrido), así que se impide la independencia. ■

Regla de multiplicación para $P(A \cap B)$

Con frecuencia la naturaleza de un experimento sugiere que dos eventos A y B deben ser supuestos independientes. Este es el caso, por ejemplo, si un fabricante recibe una tarjeta de circuito de cada uno de dos proveedores diferentes, cada tarjeta se somete a prueba al llegar y $A = \{\text{la primera está defectuosa}\}$ y $B = \{\text{la segunda está defectuosa}\}$. Si $P(A) = .1$, también deberá ser el caso de que $P(A \mid B) = .1$; sabiendo la condición de la segunda tarjeta no informa sobre la condición de la primera. La probabilidad de que ambos eventos ocurran se calcula fácilmente a partir de la probabilidad individual de los eventos cuando éstos son independientes.

PROPOSICIÓN

A y B son independientes si y sólo si

$$P(A \cap B) = P(A) \cdot P(B) \tag{2.8}$$

La verificación de esta regla de multiplicación es como sigue:

$$P(A \cap B) = P(A \mid B) \cdot P(B) = P(A) \cdot P(B) \tag{2.9}$$

donde la segunda igualdad en la ecuación (2.9) es válida si y sólo si A y B son independientes. Debido a la equivalencia de independencia y la ecuación (2.8), la segunda puede ser utilizada como definición de independencia.

Ejemplo 2.34 Se sabe que 30% de las lavadoras de cierta compañía requieren servicio mientras se encuentran dentro de garantía, en tanto que sólo 10% de sus secadoras necesitan dicho servicio. Si alguien adquiere tanto una lavadora como una secadora fabricadas por esta compañía, ¿cuál es la probabilidad de que ambas máquinas requieran servicio de garantía?

Sea A el evento en que la lavadora necesita servicio mientras se encuentra dentro de garantía y defina B de forma análoga para la secadora. Entonces $P(A) = .30$ y $P(B) = .10$. Suponiendo que las dos máquinas funcionan independientemente una de otra, la probabilidad deseada es

$$P(A \cap B) = P(A) \cdot P(B) = (.30)(.10) = .03$$ ■

Es fácil demostrar que A y B son independientes si y sólo si A' y B son independientes, A y B' son independientes y A' y B' son independientes. Por lo tanto, en el ejemplo 2.34, la probabilidad de que ninguna máquina necesite servicio es

$$P(A' \cap B') = P(A') \cdot P(B') = (.70)(.90) = .63$$

Ejemplo 2.35 Cada día, de lunes a viernes, un lote de componentes enviado por un primer proveedor arriba a una instalación de inspección. Dos días a la semana, también arriba un lote de un segundo proveedor. Ochenta por ciento de todos los lotes del proveedor 1 son inspeccionados y 90% de los del proveedor 2 también lo son. ¿Cuál es la probabilidad de que, en un día seleccionado al azar, dos lotes sean inspeccionados? Esta pregunta se responderá suponiendo que en los días en que se inspeccionan dos lotes, si el primer lote pasa es independiente de si el segundo también lo hace. La figura 2.13 muestra la información relevante.

$$\begin{aligned} P(\text{dos pasan}) &= P(\text{dos recibidos} \cap \text{ambos pasan}) \\ &= P(\text{ambos pasan} \mid \text{dos recibidos}) \cdot P(\text{dos recibidos}) \\ &= [(.8)(.9)](.4) = .288 \end{aligned}$$ ■

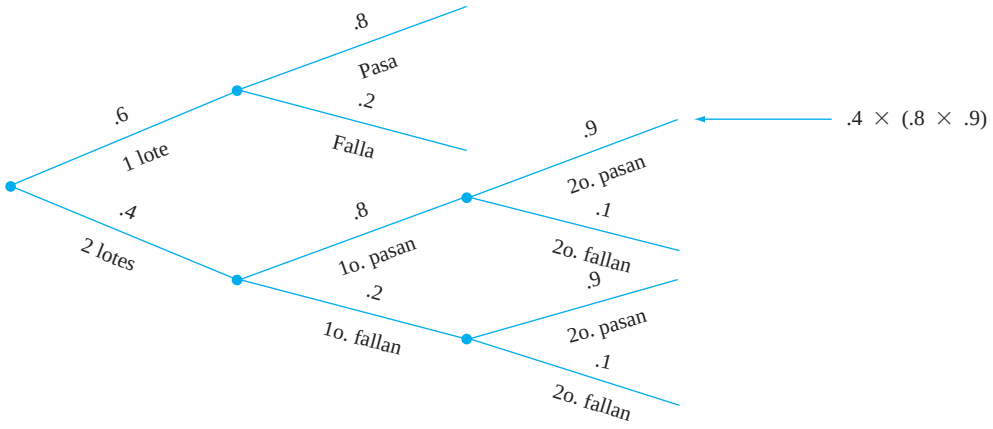


Figura 2.13 Diagrama de árbol para el ejemplo 2.35

Independencia de más de dos eventos

La noción de independencia de dos eventos puede ser ampliada a conjuntos de más de dos eventos. Aunque es posible ampliar la definición para dos eventos independientes trabajando en función de probabilidades condicionales y no condicionales, es más directo y menos tedioso seguir las líneas de la última proposición.

DEFINICIÓN

Los eventos $A_1, \ldots, A_n$ son **mutuamente independientes** si por cada k ($k = 2, 3, \ldots, n$) y cada subconjunto de índices $i_1, i_2, \ldots, i_k$,

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdot \cdots \cdot P(A_{i_k})$$

Parafraseando la definición, los eventos son mutuamente independientes si la probabilidad de la intersección de cualquier subconjunto de los n eventos es igual al producto de las probabilidades individuales. Al utilizar la propiedad de multiplicación para más de dos eventos independientes, es legítimo reemplazar una o más de las A_i por sus complementos (p. ej., si A_1, A_2 y A_3 son eventos independientes, también lo son A_1', A_2' y A_3'). Como fue el caso con dos eventos, con frecuencia se especifica al principio de un problema la independencia de ciertos eventos. La probabilidad de una intersección puede entonces ser calculada vía multiplicación.

Ejemplo 2.36 El artículo "Reliability Evaluation of Solar Photovoltaic Arrays" (*Solar Energy*, 2002: 129–141) presenta varias configuraciones de redes fotovoltaicas solares compuestas de celdas solares de silicio cristalino. Considérese primero el sistema ilustrado en la figura 2.14(a).

Existen dos subsistemas conectados en paralelo, y cada uno contiene tres celdas. Para que el sistema funcione, por lo menos uno de los dos subsistemas en paralelo debe funcio-

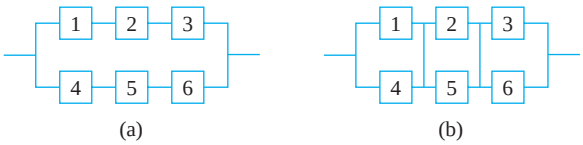


Figura 2.14 Configuraciones de sistema para el ejemplo 2.36: (a) serie-paralelo; (b) total-transversal-ligada

nar. Dentro de cada subsistema, las tres celdas están conectadas en serie, así que un subsistema funcionará sólo si todas sus celdas funcionan. Considere un valor de duración particular t_0 y suponga que desea determinar la probabilidad de que la duración del sistema exceda de t_0. Sea A_i el evento en que la duración de la celda i excede de t_0 ($i = 1, 2, \ldots, 6$). Se supone que las A_i' son eventos independientes (ya sea que cualquier celda particular que dure más de t_0 horas no tenga ningún efecto en sí o no cualquier otra celda lo hace) y que $P(A_i) = .9$ por cada i, puesto que las celdas son idénticas. Entonces

$$\begin{aligned} P(\text{la duración del sistema excede de } t_0) &= P[(A_1 \cap A_2 \cap A_3) \cup (A_4 \cap A_5 \cap A_6)] \\ &= P(A_1 \cap A_2 \cap A_3) + P(A_4 \cap A_5 \cap A_6) \\ &\quad - P[(A_1 \cap A_2 \cap A_3) \cap (A_4 \cap A_5 \cap A_6)] \\ &= (.9)(.9)(.9) + (.9)(.9)(.9) - (.9)(.9)(.9)(.9)(.9)(.9) = .927 \end{aligned}$$

Alternativamente,

$$\begin{aligned} P(\text{la duración del sistema excede de } t_0) &= 1 - P(\text{ambas duraciones del subsistema son} \le t_0) \\ &= 1 - [P(\text{la duración del subsistema es} \le t_0)]^2 \\ &= 1 - [1 - P(\text{la duración del subsistema es} > t_0)]^2 \\ &= 1 - [1 - (.9)^3]^2 = .927 \end{aligned}$$

Considérese a continuación el sistema vinculado en cruz mostrado en la figura 2.14(b), obtenido a partir de la red conectada en serie-paralelo mediante la conexión de enlaces a través de cada columna de uniones. Ahora el sistema falla en cuanto toda una columna falla, y la duración del sistema excede de t_0 sólo si la duración de cada columna lo hace. Para esta configuración,

$$\begin{aligned} P(\text{la duración del sistema es de por lo menos } t_0) &= [P(\text{la duración de la columna excede de } t_0)]^3 \\ &= [1 - P(\text{duración de la columna es} \le t_0)]^3 \\ &= [1 - P(\text{la duración de ambas celdas en una columna es} \le t_0)]^3 \\ &= [1 - (1 - .9)^2]^3 = .970 \end{aligned}$$

■

EJERCICIOS Sección 2.5 (70–89)

70. Reconsidere el escenario de la tarjeta de crédito del ejercicio 47 (sección 2.4) y demuestre que A y B son dependientes utilizando primero la definición de independencia y luego verificando que la propiedad de multiplicación no prevalece.

71. Una compañía de exploración petrolera en la actualidad tiene dos proyectos activos, uno en Asia y el otro en Europa. Sea A el evento en que el proyecto asiático tiene éxito y B el evento en que el proyecto europeo tiene éxito. Suponga que A y B son eventos independientes con $P(A) = .4$ y $P(B) = .7$.

- **a.** Si el proyecto asiático no tiene éxito, ¿cuál es la probabilidad de que el europeo tampoco tenga éxito? Explique su razonamiento.
- **b.** ¿Cuál es la probabilidad de que por lo menos uno de los dos proyectos tenga éxito?
- **c.** Dado que por lo menos uno de los dos proyectos tiene éxito, ¿cuál es la probabilidad de que sólo el proyecto asiático tenga éxito?

72. En el ejercicio 13, ¿es cualquier A_i independiente de cualquier otro A_j? Responda utilizando la propiedad de multiplicación para eventos independientes.

73. Si A y B son eventos independientes, demuestre que A' y B también son independientes. [*Sugerencia*: primero establezca una relación entre $P(A' \cap B)$, $P(B)$ y $P(A \cap B)$.]

74. Suponga que las proporciones de fenotipos sanguíneos en una población son las siguientes:

A	B	AB	O
.40	.11	.04	.45

Suponiendo que los fenotipos de dos individuos seleccionados al azar son independientes uno de otro, ¿cuál es la probabilidad de que ambos fenotipos sean O? ¿Cuál es la probabilidad de que los fenotipos de dos individuos seleccionados al azar coincidan?

75. Una de las suposiciones que sustentan la teoría de las gráficas de control (véase el capítulo 16) es que los puntos graficados sucesivos son independientes entre sí. Cada punto puede señalar que un proceso de producción está funcionando correctamente o que existe algún funcionamiento defectuoso. Aun cuando un proceso esté funcionando de manera correcta, existe

una pequeña probabilidad de que un punto particular señalará un problema con el proceso. Suponga que esta probabilidad es de .05. ¿Cuál es la probabilidad de que por lo menos uno de 10 puntos sucesivos indique un problema cuando de hecho el proceso está operando correctamente? Responda esta pregunta para 25 puntos sucesivos.

76. En octubre de 1994 se descubrió un defecto en un determinado chip Pentium instalado en las computadoras que podía dar lugar a una respuesta equivocada al realizar una división. El fabricante sostuvo inicialmente que la posibilidad de que cualquier división particular fuera incorrecta era sólo 1 de cada 9 mil millones, así que tomaría miles de años antes de que un usuario típico detectara un error. Sin embargo, los estadísticos no son usuarios típicos, algunas técnicas estadísticas modernas son tan computacionalmente intensivas que mil millones de divisiones en un corto periodo de tiempo no están fuera del reino de la posibilidad. Suponiendo que la cifra de 1 en 9 mil millones es correcta y que los resultados de las distintas divisiones son independientes el uno del otro, ¿cuál es la probabilidad de que al menos haya un error en mil millones de divisiones con este chip?

77. La costura de un avión requiere 25 remaches. La costura tendrá que ser retrabajada si alguno de los remaches está defectuoso. Suponga que los remaches están defectuosos independientemente uno de otro, cada uno con la misma probabilidad.

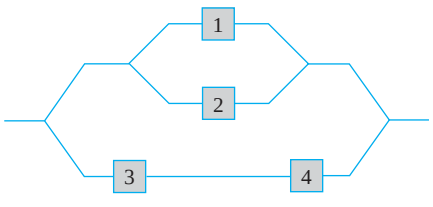
a. Si 20% de todas las costuras tienen que ser retrabajadas, ¿cuál es la probabilidad de que un remache esté defectuoso?

b. ¿Qué tan pequeña deberá ser la probabilidad de un remache defectuoso para garantizar que sólo 10% de las costuras tengan que ser retrabajadas?

78. Una caldera tiene cinco válvulas de alivio idénticas. La probabilidad de que cualquier válvula particular se abra en un momento de demanda es de .95. Suponiendo que operan independientemente, calcule P(por lo menos una válvula se abre) y P(por lo menos una válvula no se abre).

79. Dos bombas conectadas en paralelo fallan independientemente una de otra en cualquier día dado. La probabilidad de que falle sólo la bomba más vieja es de .10 y la probabilidad de que sólo la bomba más nueva falle es de .05. ¿Cuál es la probabilidad de que el sistema de bombeo falle en cualquier día dado (lo que sucede si ambas bombas fallan)?

80. Considere el sistema de componentes conectados como en la figura adjunta. Los componentes 1 y 2 están conectados en paralelo, de modo que el subsistema trabaja si y sólo si 1 o 2 trabaja; como 3 y 4 están conectados en serie, ¿el subsistema trabaja si y sólo si 3 y 4 trabajan? Si los componentes funcionan independientemente uno de otro y P(el componente trabaja) = .9, calcule P(el sistema trabaja).



81. Remítase otra vez al sistema en serie-paralelo introducido en el ejemplo 2.35 y suponga que existen sólo dos celdas en lugar de tres en cada subsistema en paralelo [en la figura 2.14(a), elimine las celdas 3 y 6 y renumere las celdas 4 y 5 como 3 y 4]. Utilizando $P(A_i) = .9$, es fácil ver que la probabilidad de que la duración del sistema exceda de t_0 es de .9639. ¿A qué valor tendría que cambiar .9 para incrementar la confiabilidad y duración del sistema de .9639 a .99? [*Sugerencia*: sea $P(A_i) = p$, exprese la confiabilidad del sistema en función de p, luego haga $x = p^2$.]

82. Considere lanzar en forma independiente dos dados imparciales, uno rojo y otro verde. Sea A el evento en que el dado rojo muestra 3 puntos, B el evento en que el dado verde muestra 4 puntos y C el evento en que el número total de puntos que muestran los dos dados es 7. ¿Son estos eventos independientes por pares (es decir, ¿son A y B eventos independientes, son A y C independientes y son B y C independientes? ¿Son los tres eventos mutuamente independientes?

83. Los componentes enviados a un distribuidor son revisados en cuanto a defectos por dos inspectores diferentes (cada componente es revisado por ambos inspectores). El primero detecta 90% de todos los defectuosos que están presentes y el segundo hace lo mismo. Por lo menos un inspector no detecta un defecto en 20% de todos los componentes defectuosos. ¿Cuál es la probabilidad de que ocurra lo siguiente?

a. ¿Un componente defectuoso será detectado sólo por el primer inspector? ¿Por exactamente uno de los dos inspectores?

b. ¿Los tres componentes defectuosos en un lote no son detectados por ambos inspectores (suponiendo que las inspecciones de los diferentes componentes son independientes unas de otras)?

84. Setenta por ciento de todos los vehículos examinados en un centro de verificación de emisiones pasan la inspección. Suponiendo que vehículos sucesivos pasan o fallan independientemente uno de otro, calcule las siguientes probabilidades:

a. P(los tres vehículos siguientes inspeccionados pasan)

b. P(por lo menos uno de los tres vehículos siguientes pasa)

c. P(exactamente uno de los tres vehículos siguientes pasa)

d. P(cuando mucho uno de los tres vehículos siguientes inspeccionados pasa)

e. Dado que por lo menos uno de los tres vehículos siguientes pasa la inspección, ¿cuál es la probabilidad de que los tres pasen (una probabilidad condicional)?

85. Un inspector de control de calidad verifica artículos recién producidos en busca de fallas. El inspector examina un artículo en busca de fallas en una serie de observaciones independientes, cada una de duración fija. Dado que en realidad está presente una imperfección, sea p la probabilidad de que la imperfección sea detectada durante cualquier observación (este modelo se discute en "Human Performance in Sampling Inspection", *Human Factors*, 1979: 99–105).

a. Suponiendo que un artículo tiene una imperfección, ¿cuál es la probabilidad de que sea detectada al final de la segunda observación (una vez que una imperfección ha sido detectada, la secuencia de observaciones termina)?

b. Dé una expresión para la probabilidad de que una imperfección será detectada al final de la n-ésima observación.

c. Si cuando en tres observaciones no ha sido detectada una imperfección, el artículo es aprobado, ¿cuál es la probabilidad de que un artículo imperfecto pase la inspección?

d. Suponga que 10% de todos los artículos contienen una imperfección [P(artículo seleccionado al azar muestra una imperfección) = .1]. Con la suposición del inciso (c), ¿cuál

es la probabilidad de que un artículo seleccionado al azar pase la inspección (pasará automáticamente si no muestra una imperfección, pero también podría pasar si muestra una imperfección)?

e. Dado que un artículo ha pasado la inspección (ninguna imperfección en tres observaciones), ¿cuál es la probabilidad de que sí tenga una imperfección? Calcule para $p = .5$.

86. a. Una compañía maderera acaba de recibir un lote de 10,000 tablas de 2×4. Suponga que 20% de estas tablas (2,000) en realidad están demasiado tiernas o verdes para ser utilizadas en construcción de primera calidad. Se eligen dos tablas al azar, una después de la otra. *Sea* $A = \{$la primera tabla está verde$\}$ y $B = \{$la segunda tabla está verde$\}$. Calcule $P(A)$, $P(B)$ y $P(A \cap B)$ (un diagrama de árbol podría ayudar). ¿Son A y B independientes?

b. Con A y B independientes y $P(A) = P(B) = .2$, ¿cuál es $P(A \cap B)$? ¿Cuánta diferencia existe entre esta respuesta y $P(A \cap B)$ en el inciso (a)? Para propósitos de cálculo $P(A \cap B)$, ¿se puede suponer que A y B del inciso (a) son independientes para obtener en esencia la probabilidad correcta?

c. Suponga que un lote consta de 10 tablas, de las cuales dos están verdes. ¿Produce ahora la suposición de independencia aproximadamente la respuesta correcta para $P(A \cap B)$? ¿Cuál es la diferencia crítica entre la situación en este caso y la del inciso (a)? ¿Cuándo piensa que una suposición de independencia sería válida al obtener una respuesta aproximadamente correcta para $P(A \cap B)$?

87. Considere la posibilidad de seleccionar al azar una sola persona y que ésta prueba tres vehículos diferentes. Defina los eventos A_1, A_2 y A_3 por

$A_1 =$ como el vehículo #1 $\qquad A_2 =$ como el vehículo #2

$A_3 =$ como el vehículo #3

Suponga que $P(A_1) = .55$, $P(A_2) = .65$, $P(A_3) = .70$, $P(A_1 \cup A_2) = .80$, $P(A_2 \cap A_3) = .40$ y $P(A_1 \cup A_2 \cup A_3) = .88$.

a. ¿Cuál es la probabilidad de que a la persona le guste tanto el vehículo # 1 como el vehículo # 2?

b. Determinar e interpretar $P(A_2|A_3)$.

c. ¿Son eventos independientes A_2 y A_3? Responda en dos maneras diferentes.

d. Si usted se entera de que a la persona no le gusta el vehículo # 1, ¿cuál es ahora la probabilidad de que a él/ella le guste por lo menos uno de los otros dos vehículos?

88. El profesor Stan der Deviation puede tomar una de dos rutas en el trayecto del trabajo a su casa. En la primera ruta, hay cuatro cruces de ferrocarril. La probabilidad de que sea detenido por un tren en cualquiera de los cruces es .1 y los trenes operan independientemente en los cuatro cruces. La otra ruta es más larga pero sólo hay dos cruces, independientes uno de otro, con la misma posibilidad de que sea detenido por un tren al igual que en la primera ruta. En un día particular, el profesor Deviation tiene una reunión programada en casa durante cierto tiempo. ¿Cualquier ruta que tome, calcula que llegará tarde si es detenido por los trenes en por lo menos la mitad de los cruces encontrados.

a. ¿Cuál ruta deberá tomar para reducir al mínimo la probabilidad de llegar tarde a la reunión?

b. Si lanza al aire una moneda imparcial para decidir qué ruta tomar y está retrasado, ¿cuál es la probabilidad de que tome la ruta de los cuatro cruces?

89. Suponga que se colocan etiquetas idénticas en las dos orejas de un zorro. El zorro es dejado en libertad durante un lapso de tiempo. Considere los dos eventos $C_1 = \{$se pierde la etiqueta de la oreja izquierda$\}$ y $C_2 = \{$se pierde la etiqueta de la oreja derecha$\}$. Sea $\pi = P(C_1) = P(C_2)$ y suponga que C_1 y C_2 son eventos independientes. Deduzca una expresión (que implique π) para la probabilidad de que exactamente una etiqueta se pierda dado que cuando mucho una se pierde ("Ear Tag Loss in Red Foxes", *J. Wildlife Mgmt.*, 1976: 164–167). [*Sugerencia*: trace un diagrama de árbol en el cual las dos ramas iniciales se refieren a si la etiqueta de la oreja izquierda se pierde.]

EJERCICIOS SUPLEMENTARIOS (90–114)

90. Una pequeña compañía manufacturera va a echar a andar un turno de noche. Hay 20 mecánicos empleados por la compañía.

a. Si una cuadrilla nocturna se compone de 3 mecánicos, ¿cuántas cuadrillas diferentes son posibles?

b. Si los mecánicos están clasificados 1, 2, . . . , 20 en orden de competencia, ¿cuántas de estas cuadrillas no incluirían al mejor mecánico?

c. ¿Cuántas de las cuadrillas tendrían por lo menos 1 de los 10 mejores mecánicos?

d. Si se selecciona al azar una de estas cuadrillas para que trabajen una noche particular, ¿cuál es la probabilidad de que el mejor mecánico no trabaje esa noche?

91. Una fábrica utiliza tres líneas de producción para fabricar latas de cierto tipo. La tabla adjunta da porcentajes de latas que no cumplen con las especificaciones, clasificadas por tipo de incumplimiento de las especificaciones, para cada una de las tres líneas durante un lapso de tiempo particular.

	Línea 1	Línea 2	Línea 3
Manchas	15	12	20
Grietas	50	44	40
Problemas con la argolla	21	28	24
Defecto superficial	10	8	15
Otro	4	8	2

Durante este periodo, la línea 1 produjo 500 latas fuera de especificación, la 2 produjo 400 latas como ésas y la 3 fue responsable de 600 latas fuera de especificación. Suponga que se selecciona al azar una de estas 1500 latas.

a. ¿Cuál es la probabilidad de que la lata venga de la línea 1? ¿Cuál es la probabilidad de que la razón del incumplimiento de la especificación sea una grieta?

b. Si la lata seleccionada provino de la línea 1, ¿cuál es la probabilidad de que tenga una mancha?

c. Dado que la lata seleccionada mostró un defecto superficial, ¿cuál es la probabilidad de que provenga de la línea 1?

92. Un empleado de la oficina de inscripciones en una universidad en este momento tiene diez formas en su escritorio en espera de ser procesadas. Seis de éstas son peticiones de baja y las otras cuatro son solicitudes de sustitución de curso.

a. Si selecciona al azar seis de estas formas para dárselas a un subordinado, ¿cuál es la probabilidad de que sólo uno de los dos tipos de formas permanezca en su escritorio?

b. Suponga que tiene tiempo para procesar sólo cuatro de estas formas antes de salir del trabajo. Si estas cuatro se seleccionan al azar una por una, ¿cuál es la probabilidad de que cada forma subsiguiente sea de un tipo diferente de su predecesora?

93. Un satélite está programado para ser lanzado desde Cabo Cañaveral, en Florida, y otro lanzamiento está programado para la Base de la Fuerza Aérea Vandenberg en California. Sea A el evento en que el lanzamiento en Vandenberg se hace a la hora programada y B el evento en que el lanzamiento en Cabo Cañaveral se hace a la hora programada. Si A y B son eventos independientes con $P(A) > P(B)$, $P(A \cup B) = .626$ y $P(A \cap B) = .144$, determine los valores de $P(A)$ y $P(B)$.

94. Un transmisor envía un mensaje utilizando un código binario, esto es, una secuencia de ceros y unos. Cada bit transmitido (0 o 1) debe pasar a través de tres relevadores para llegar al receptor. En cada relevador, la probabilidad es .20 de que el bit enviado será diferente del bit recibido (una inversión). Suponga que los relevadores operan independientemente uno de otro.

transmisor $\rightarrow$ relevador 1 $\rightarrow$ relevador 2 $\rightarrow$ relevador 3 $\rightarrow$ receptor

a. Si el transmisor envía un 1, ¿cuál es la probabilidad de que los tres relevadores envíen un 1?

b. Si el transmisor envía un 1, ¿cuál es la probabilidad de que el receptor reciba un 1? [*Sugerencia*: los ocho resultados experimentales pueden ser mostrados en un diagrama de árbol con tres ramas de generación, una por cada relevador.]

c. Suponga que 70% de todos los bits enviados por el transmisor son unos. Si el receptor recibe un 1, ¿cuál es la probabilidad de que un 1 haya sido enviado?

95. El individuo A tiene un círculo de cinco amigos cercanos (B, C, D, E y F). A escuchó cierto rumor originado fuera del círculo e invitó a sus cinco amigos a una fiesta para contarles el rumor. Para empezar, A escoge a uno de los cinco al azar y se lo cuenta. Dicho individuo escoge entonces al azar a uno de los cuatro individuos restantes y repite el rumor. Después, de aquellos que ya oyeron el rumor uno se lo cuenta a otro nuevo individuo y así hasta que todos oyen el rumor.

a. ¿Cuál es la probabilidad de que el rumor se repita en el orden B, C, D, E y F?

b. ¿Cuál es la probabilidad de que F sea la tercera persona en la reunión a la que se le contará el rumor?

c. ¿Cuál es la probabilidad de que F sea la última persona en oír el rumor?

d. Si en cada etapa la persona que en ese momento "tiene" el rumor no sabe quien ya lo ha escuchado y selecciona al siguiente destinatario aleatoriamente de entre cinco individuos posibles, ¿cuál es la probabilidad de que F no haya escuchado todavía el rumor después de haber sido dicho 10 veces en la fiesta?

96. De acuerdo con el artículo "Optimization of Distribution Parameters for Estimating Probability of Crack Detection" (*J. of Aircraft*, 2009: 2090–2097), la siguiente ecuación de "Palmberg" se usa comúnmente para determinar la probabilidad $P_d(c)$ de la detección de una grieta de tamaño c en la estructura de la aeronave:

$$P_d(c) = \frac{(c/c^*)^\beta}{1 + (c/c^*)^\beta}$$

donde c^* es el tamaño de la grieta que corresponde a una probabilidad de detección de .5 (y por tanto es una evaluación de la calidad del proceso de inspección).

a. Compruebe que $P_d(c^*) = .5$

b. ¿Qué es $P_d(2c^*)$ cuando $\beta = 4$?

c. Supongamos que un inspector revisa dos paneles diferentes, uno con un tamaño de grieta de c^* y otro con un tamaño de grieta de $2c^*$. Una vez más, suponiendo $\beta = 4$ y también que los resultados de las dos inspecciones son independientes el uno del otro, ¿cuál es la probabilidad de que exactamente una de las dos grietas se detecte?

d. ¿Qué le sucede a $P_d(c)$ cuando $\beta \rightarrow \infty$?

97. Un ingeniero químico está interesado en determinar si cierta impureza está presente en un producto. Un experimento tiene una probabilidad de .80 de detectarla si está presente. La probabilidad de no detectarla si está ausente es de .90. Las probabilidades previas de que la impureza esté presente o ausente son de .40 y .60, respectivamente. Tres experimentos distintos producen sólo dos detecciones. ¿Cuál es la probabilidad posterior de que la impureza esté presente?

98. A cada concursante en un programa de preguntas se le pide que especifique una de seis posibles categorías de entre las cuales se le hará una pregunta. Suponga P(el concursante escoge la categoría i) $= \frac{1}{6}$ y concursantes sucesivos escogen sus categorías independientemente uno del otro. Si participan tres concursantes en cada programa y los tres en un programa particular seleccionan diferentes categorías, ¿cuál es la probabilidad de que exactamente uno seleccione la categoría 1?

99. Los sujetadores roscados utilizados en la fabricación de aviones son levemente doblados para que queden bien apretados y no se aflojen durante vibraciones. Suponga que 95% de todos los sujetadores pasan una inspección inicial. De 5% que fallan, 20% están tan seriamente defectuosos que deben ser desechados. Los sujetadores restantes son enviados a una operación de redoblado, donde 40% no pueden ser recuperados y son desechados. El otro 60% de estos sujetadores son corregidos por el proceso de redoblado y posteriormente pasan la inspección.

a. ¿Cuál es la probabilidad de que un sujetador que acaba de llegar seleccionado al azar pase la inspección inicialmente o después del redoblado?

b. Dado que un sujetador pasó la inspección, ¿cuál es la probabilidad de que apruebe la inspección inicial y de que no necesite redoblado?

100. Un porcentaje de todos los individuos en una población son portadores de una enfermedad particular. Una prueba de diagnóstico para esta enfermedad tiene una tasa de detección de 90% para portadores y de 5% para no portadores. Suponga que la prueba se aplica independientemente a dos muestras de sangre diferentes del mismo individuo seleccionado al azar.

a. ¿Cuál es la probabilidad de que ambas pruebas den el mismo resultado?

b. Si ambas pruebas son positivas, ¿cuál es la probabilidad de que el individuo seleccionado sea un portador?

101. Un sistema consta de dos componentes. La probabilidad de que el segundo componente funcione de manera satisfactoria durante su duración de diseño es de .9, la probabilidad de que por lo menos uno de los dos componentes lo haga es de .96 y la probabilidad de que ambos componentes lo hagan es de .75. Dado que el primer componente funciona de manera satisfactoria durante toda su duración de diseño, ¿cuál es la probabilidad de que el segundo también lo haga?

102. Cierta compañía envía 40% de sus paquetes de correspondencia nocturna vía un servicio de correo exprés E_1. De estos paquetes, 2% llegan después del tiempo de entrega garantizado (sea L el evento "entrega demorada"). Si se selecciona al azar un registro de correspondencia nocturna del archivo de la compañía, ¿cuál es la probabilidad de que el paquete se fue vía E_1 y llegó demorado?

103. Remítase al ejercicio 102. Suponga que 50% de los paquetes nocturnos se envían vía el servicio de correo exprés E_2 y el 10% restante se envía vía E_3. De los paquetes enviados vía E_2, sólo 1% llegaron demorados, en tanto que 5% de los paquetes manejados por E_3 llegaron demorados.

a. ¿Cuál es la probabilidad de que un paquete seleccionado al azar llegue demorado?

b. Si un paquete seleccionado al azar llegó a tiempo, ¿cuál es la probabilidad de que no haya sido enviado vía E_1?

104. Una compañía utiliza tres líneas de ensamble diferentes: A_1, A_2 y A_3 para fabricar un componente particular. De los fabricados por la línea A_1, 5% tienen que ser retrabajados para corregir un defecto, mientras que 8% de los componentes de A_2 tienen que ser retrabajados y 10% de los componentes de A_3 tienen que ser retrabajados. Suponga que 50% de todos los componentes los produce la línea A_1, 30% la línea A_2 y 20% la línea A_3. Si un componente seleccionado al azar tiene que ser retrabajado, ¿cuál es la probabilidad de que provenga de la línea A_1? ¿De la línea A_2? ¿De la línea A_3?

105. Desechando la posibilidad de cumplir años el 29 de febrero, suponga que es igualmente probable que un individuo seleccionado al azar haya nacido en cualquiera de los demás 365 días.

a. Si se seleccionan al azar diez personas, ¿cuál es la probabilidad de que tengan diferentes cumpleaños? ¿De que por lo menos dos tengan el mismo cumpleaños?

b. Si k reemplaza a diez en el inciso (a), ¿cuál es la k más pequeña para la cual existe por lo menos una probabilidad de 50-50 de que dos o más personas tengan el mismo cumpleaños?

c. Si se seleccionan diez personas al azar, ¿cuál es la probabilidad de que por lo menos dos tengan el mismo cumpleaños o por lo menos dos tengan los mismos tres últimos dígitos de sus números del Seguro Social? [*Nota*: el artículo "Methods for Studying Coincidences" (F. Mosteller y P. Diaconis, *J. Amer. Stat. Assoc.*, 1989: 853–861) discute problemas de este tipo.]

106. Un método utilizado para distinguir entre rocas graníticas (G) y basálticas (B) es examinar una parte del espectro infrarrojo de la energía solar reflejada por la superficie de la roca. Sean R_1, R_2 y R_3 intensidades espectrales medidas a tres longitudes de onda diferentes; en general, para granito $R_1 < R_2 < R_3$, en tanto que para basalto $R_3 < R_1 < R_2$. Cuando se hacen mediciones a distancia (mediante un avión), varios ordenamientos de R_i pueden presentarse ya sea que la roca sea basalto o granito. Vuelos sobre regiones de composición conocida han arrojado la siguiente información:

	Graníticas	**Basálticas**
$R_1 < R_2 < R_3$	60%	10%
$R_1 < R_3 < R_2$	25%	20%
$R_3 < R_1 < R_2$	15%	70%

Suponga que para una roca seleccionada al azar en cierta región, $P(\text{granito}) = .25$ y $P(\text{basalto}) = .75$.

a. Demuestre que $P(\text{granito} \mid R_1 < R_2 < R_3) > P(\text{basalto} \mid R_1 < R_2 < R_3)$. Si las mediciones dieron $R_1 < R_2 < R_3$, ¿clasificaría la roca como granito o como basalto?

b. Si las mediciones dieron $R_1 < R_3 < R_2$, ¿cómo clasificaría la roca? Responda la misma pregunta para $R_3 < R_1 < R_2$.

c. Con las reglas de clasificación indicadas en los incisos (a) y (b) cuando se selecciona una roca de esta región, ¿cuál es la probabilidad de una clasificación errónea? [*Sugerencia*: G podría ser clasificada como B o B como G y $P(B)$ y $P(G)$ son conocidas.]

d. Si $P(\text{granito}) = p$ en lugar de .25, ¿existen valores de p (aparte de 1) para los cuales una roca siempre sería clasificada como granito?

107. A un sujeto se le permite una secuencia de vistazos para detectar un objetivo. Sea $G_i = \{\text{el objetivo es detectado en el vistazo } i\text{-ésimo}\}$, con $p_i = P(G_i)$. Suponga que los G_i son eventos independientes y escriba una expresión para la probabilidad de que el objetivo haya sido detectado al final del vistazo n-ésimo. [*Nota*: este modelo se discute en "Predicting Aircraft Detectability", *Human Factors*, 1979: 277–291.]

108. En un juego de beisbol de ligas pequeñas, el lanzador del equipo A lanza un "strike" 50% del tiempo y una bola 50% del tiempo; los lanzamientos sucesivos son independientes unos de otros y el lanzador nunca golpea a un bateador. Sabiendo esto, el mánager del equipo B ha instruido al primer bateador para que no le batee a nada.

Calcule la probabilidad de que

a. El bateador reciba base por bolas en el cuarto lanzamiento

b. El bateador reciba base por bolas en el sexto lanzamiento (por lo que dos de los primeros cinco deben ser "strikes"), por medio de un argumento de conteo o un diagrama de árbol.

c. El bateador recibe base por bolas.

d. El primer bateador anota mientras no hay ningún "out" (suponiendo que cada bateador utiliza la estrategia de no batearle a nada)

109. Cuatro ingenieros, A, B, C y D han sido citados para entrevistas de trabajo a las 10 a.m. el viernes 13 de enero, en Random Sampling, Inc. El gerente de personal ha programado a los cuatro para las oficinas de entrevistas 1, 2, 3 y 4, respectivamente. Sin embargo, el secretario del gerente no está enterado de esto, por lo que los asigna a las oficinas de un modo completamente aleatorio (¡qué más!) ¿Cuál es la probabilidad de que

a. los cuatro terminen en la oficina correcta?

b. ninguno de los cuatro termine en la oficina correcta?

110. Una aerolínea particular opera vuelos a las 10 a.m. de Chicago a Nueva York, Atlanta y Los Ángeles. Sea A el evento en que el vuelo a Nueva York está lleno y defina los eventos B y C en forma análoga para los otros dos vuelos. Suponga que $P(A) = .6$, $P(B) = .5$, $P(C) = .4$ y los tres eventos son independientes. ¿Cuál es la probabilidad de que

a. los tres vuelos estén llenos? ¿Que por lo menos uno no esté lleno?

b. sólo el vuelo a Nueva York esté lleno? ¿Que exactamente uno de los tres vuelos esté lleno?

111. Un gerente de personal va a entrevistar a cuatro candidatos para un puesto. Éstos están clasificados como 1, 2, 3 y 4 en orden de preferencia, y serán entrevistados en orden aleatorio. Sin embargo, al final de cada entrevista, el gerente sabrá sólo cómo se compara el candidato actual con los candidatos previamente entrevistados. Por ejemplo, el orden de entrevista 3, 4, 1, 2 no genera información después de la primera entrevista, muestra que el segundo candidato es peor que el primero y que el tercero es mejor que los primeros dos. Sin embargo, el orden 3, 4, 2, 1 generaría la misma información después de cada una de las primeras tres entrevistas. El gerente desea contratar al mejor candidato pero debe tomar una decisión irrevocable de contratarlo o no contratarlo después de cada entrevista. Considere la siguiente estrategia: rechazar automáticamente al primer candidato s y luego contratar al primer candidato subsiguiente que resulte mejor entre los que ya fueron entrevistados (si tal candidato no aparece, el último entrevistado es el contratado).

Por ejemplo, con $s = 2$, el orden 3, 4, 1, 2 permitiría contratar al mejor, en tanto que el orden 3, 1, 2, 4, no. De los cuatro valores s posibles (0, 1, 2 y 3), ¿cuál incrementa al máximo a P(el mejor es contratado)? [*Sugerencia*: escriba los 24 ordenamientos de entrevista igualmente probables: $s = 0$ significa que el primer candidato es automáticamente contratado.]

112. Considere cuatro eventos independientes A_1, A_2, A_3 y A_4, y sea $p_i = P(A_i)$ con $i = 1, 2, 3, 4$. Exprese la probabilidad de que por lo menos uno de estos eventos ocurra en función de las p_i y haga lo mismo para la probabilidad de que por lo menos dos de los eventos ocurran.

113. Una caja contiene los siguientes cuatro papelitos y cada uno tiene exactamente las mismas dimensiones: (1) gana el premio 1: (2) gana el premio 2: (3) gana el premio 3; (4) gana los premios 1, 2, y 3. Se selecciona un papelito al azar. Sea $A_1 = \{\text{gana el premio 1}\}$, $A_2 = \{\text{gana el premio 2}\}$ y $A_3 = \{\text{gana el premio 3}\}$. Demuestre que A_1 y A_2 son independientes, que A_1 y A_3 son independientes, y que A_2 y A_3 también son independientes (ésta es una independencia *por pares*). Sin embargo, demuestre que $P(A_1 \cap A_2 \cap A_3) \neq P(A_1) \cdot P(A_2) \cdot P(A_3)$, así que los tres eventos *no* son mutuamente independientes.

114. Demuestre que si A_1, A_2 y A_3 son eventos independientes, entonces $P(A_1 \mid A_2 \cap A_3) = P(A_1)$.

Bibliografía

Durrett, Richard, *Elementary Probability for Applications*, Cambridge Univ. Press, Londres, Inglaterra, 2009. Una presentación concisa a un nivel un poco más alto que este texto.

Mosteller, Frederick, Robert Rourke y George Thomas, *Probability with Statistical Applications* (2a. ed.), Addison-Wesley, Reading, MA, 1970. Una muy buena introducción a la probabilidad, con muchos ejemplos entretenidos; especialmente buenos con respecto a reglas de conteo y su aplicación.

Olkin, Ingram, Cyrus Derman y Leon Gleser, *Probability Models and Application* (2a. ed.), Macmillan, Nueva York, 1994. Una amplia introducción a la probabilidad escrita a un nivel matemático un poco más alto que este texto pero que contiene muchos buenos ejemplos.

Ross, Sheldon, *A First Course in Probability* (6a. ed.), Macmillan, Nueva York, 2009. Algo concisamente escrito y más matemáticamente complejo que este texto pero contiene una gran cantidad de ejemplos y ejercicios interesantes.

Winkler, Robert, *Introduction to Bayesian Inference and Decision*, Holt, Rinehart & Winston, Nueva York, 1972. Una muy buena introducción a la probabilidad subjetiva.

3 Variables aleatorias discretas y distribuciones de probabilidad

INTRODUCCIÓN

Ya sea que un experimento produzca resultados cualitativos o cuantitativos, los métodos de análisis estadístico requieren enfocarse en ciertos aspectos numéricos de los datos (como la proporción muestral x/n, la media $\bar{x}$ o la desviación estándar σ). El concepto de variable aleatoria permite pasar de los resultados experimentales a la función numérica de los resultados. Existen dos tipos fundamentalmente diferentes de variables aleatorias: las variables aleatorias discretas y las variables aleatorias continuas. En este capítulo se examinan las propiedades básicas y se discuten los ejemplos más importantes de variables discretas. El capítulo 4 se enfoca en las variables aleatorias continuas.

3.1 Variables aleatorias

En cualquier experimento existen numerosas características que pueden ser observadas o medidas, pero en la mayoría de los casos un experimentador se enfoca en algún aspecto específico o aspectos de una muestra. Por ejemplo, en un estudio de patrones de viaje entre los suburbios y la ciudad en un área metropolitana, a cada individuo en una muestra se le podría preguntar sobre la distancia que recorre para ir de su casa al trabajo y viceversa, y el número de personas que viajan en el mismo vehículo, pero no sobre su coeficiente intelectual, ingreso, tamaño de su familia y otras características. Por otra parte, un investigador puede probar una muestra de componentes y anotar sólo el número de los que han fallado dentro de 1000 horas, en lugar de anotar los tiempos de falla individuales.

En general, cada resultado de un experimento puede ser asociado con un número especificando una regla de asociación (p. ej., el número entre la muestra de diez componentes que no duran 1000 horas o el peso total del equipaje en una muestra de 25 pasajeros de aerolínea). Semejante regla de asociación se llama **variable aleatoria**, variable porque diferentes valores numéricos son posibles y aleatoria porque el valor observado depende de cuál de los posibles resultados experimentales resulte (figura 3.1).

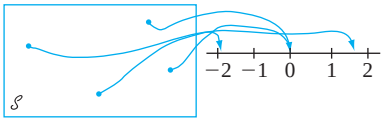


Figura 3.1 Una variable aleatoria

DEFINICIÓN

Para un espacio muestral dado $\mathcal{S}$ de algún experimento, una **variable aleatoria** es cualquier regla que asocia un número con cada resultado en $\mathcal{S}$. En lenguaje matemático, una variable aleatoria es una función cuyo dominio es el espacio muestral y cuyo rango es el conjunto de los números reales.

Se acostumbra denotar las variables aleatorias con letras mayúsculas, tales como X y Y, que son las de cerca del final del alfabeto. En contraste al uso previo de una letra minúscula, tal como x, para denotar una variable, ahora se utilizarán letras minúsculas para representar algún valor particular de la variable aleatoria correspondiente. La notación $X(s) = x$ significa que x es el valor asociado con el resultado s por medio de la variable aleatoria X.

Ejemplo 3.1 Cuando un estudiante llama a un servicio de asistencia universitaria para apoyo técnico, él/ella podrá inmediatamente hablar con alguien (S) o será puesto en espera (F). Con $\mathcal{S} = \{S, F\}$, la variable aleatoria X se define como

$$X(S) = 1 \quad X(F) = 0$$

La variable aleatoria X indica si (1) o no (0) el estudiante puede hablar inmediatamente con alguien. ■

La variable aleatoria X en el ejemplo 3.1 se especificó al poner en lista explícitamente cada elemento de $\mathcal{S}$ y el número asociado. Una lista como ésa es tediosa si $\mathcal{S}$ contiene más de algunos cuantos resultados, pero con frecuencia puede ser evitada.

Ejemplo 3.2 Considere el experimento en el cual se marca un número telefónico en cierto código de área con un marcador de números aleatorio (tales dispositivos los utilizan en forma extensa en organizaciones encuestadoras) y defina una variable aleatoria Y como

$$Y = \begin{cases} 1 & \text{si el número seleccionado no aparece en el directorio} \\ 0 & \text{si el número seleccionado sí aparece en el directorio} \end{cases}$$

Por ejemplo, si 5282966 aparece en el directorio telefónico, entonces $Y(5282966) = 0$ en tanto que $Y(7727350) = 1$ dice que el número 7727350 no aparece en el directorio telefónico. Una descripción en palabras de esta índole es más económica que una lista completa, por lo que se utilizará tal descripción siempre que sea posible. ■

En los ejemplos 3.1 y 3.2, los únicos valores posibles de la variable aleatoria fueron 0 y 1. Tal variable aleatoria se presenta con suficiente frecuencia como para darle un nombre especial, en honor del individuo que la estudió primero.

DEFINICIÓN

Cualquier variable aleatoria cuyos únicos valores posibles son 0 y 1 se llama **variable aleatoria de Bernoulli.**

En ocasiones se deseará definir y estudiar diferentes variables del mismo espacio muestral.

Ejemplo 3.3 El ejemplo 2.3 describe un experimento en el cual se determinó el número de bombas en uso en cada una de dos gasolineras. Defina las variables aleatorias X, Y y U como

X = el número total de bombas en uso en las dos gasolineras

Y = la diferencia entre el número de bombas en uso en la gasolinera 1 y el número en uso en la gasolinera 2

U = el máximo de los números de bombas en uso en las dos gasolineras

Si se realiza este experimento y se obtiene $s = (2, 3)$, entonces $X((2, 3)) = 2 + 3 = 5$, por lo que se dice que el valor observado de X fue $x = 5$. Asimismo, el valor observado de Y sería $y = 2 - 3 = -1$ y el de U sería $u = \text{máx}(2, 3) = 3$. ■

Cada una de las variables aleatorias de los ejemplos 3.1–3.3 puede asumir sólo un número finito de posibles valores. Éste no tiene que ser el caso.

Ejemplo 3.4 Se considera un experimento en que se examinaron baterías de 9 volts hasta que se obtuvo una con un voltaje aceptable (S). El espacio muestral es $\mathcal{S} = \{S, FS, FFS, \ldots\}$. Defina una variable aleatoria X como

X = el número de baterías examinadas antes de que se termine el experimento

En ese caso $X(S) = 1$, $X(FS) = 2$, $X(FFS) = 3, \ldots,$ $X(FFFFFS) = 7$, y así sucesivamente. Cualquier entero positivo es un valor positivo de X, así que el conjunto de valores posibles es infinito. ■

Ejemplo 3.5 Suponga que del mismo modo aleatorio se selecciona un lugar (latitud y longitud) en el territorio continental de Estados Unidos. Defina una variable aleatoria Y como

Y = la altura sobre el nivel del mar en el lugar seleccionado

Por ejemplo, si el lugar seleccionado fuera (39°50′N, 98°35′O), entonces se podría tener $Y((39°50'N, 98°35'O)) = 1748.26$ pies. El valor más grande posible de Y es 14,494 (Monte Whitney) y el valor más pequeño posible es −282 (Valle de la Muerte). El conjunto de todos los valores posibles de Y es el conjunto de todos los números en el intervalo entre −282 y 14,494, es decir,

$$\{y : y \text{ es un número}, -282 \le y \le 14{,}494\}$$

y existe un número infinito de números en este intervalo. ■

Dos tipos de variables aleatorias

En la sección 1.2, se distinguió entre los datos que resultan de observaciones de una variable de conteo y los datos obtenidos observando valores de una variable de medición. Una distinción un poco más formal caracteriza dos tipos diferentes de variables aleatorias.

DEFINICIÓN

Una variable aleatoria **discreta** es una variable aleatoria cuyos valores posibles constituyen un conjunto finito o bien pueden ser puestos en lista en una secuencia infinita en la cual existe un primer elemento, un segundo elemento, y así sucesivamente ("contablemente" infinita).

Una variable aleatoria es **continua** si *ambas* de las siguientes condiciones se cumplen:

1. Su conjunto de valores posibles se compone de todos los números que hay en un solo intervalo sobre la línea de numeración (posiblemente de extensión infinita, es decir, desde $-\infty$ hasta ∞) o todos los números en una unión disjunta de dichos intervalos (por ejemplo, $[0, 10] \cup [20, 30]$).
2. Ningún valor posible de la variable tiene probabilidad positiva, esto es, $P(X = c) = 0$ con cualquier valor posible de c.

Aunque cualquier intervalo sobre la línea de numeración contiene un número infinito de números, se puede demostrar que no existe ninguna forma de crear una lista infinita de todos estos valores, pues existen demasiados de ellos. La segunda condición que describe una variable aleatoria continua es tal vez contraintuitiva, puesto que parecería que implica una probabilidad total de cero para todos los valores posibles. Pero en el capítulo 4 se verá que los *intervalos* de valores tienen probabilidad positiva; la probabilidad de un intervalo se reducirá a cero a medida que su ancho tienda a cero.

Ejemplo 3.6 Todas las variables aleatorias de los ejemplos 3.1–3.4 son discretas. Como otro ejemplo, suponga que se eligen al azar parejas de casados y que a cada persona se le hace una prueba de sangre hasta encontrar un esposo y esposa con el mismo factor Rh. Con $X =$ el número de pruebas de sangre que serán realizadas, los posibles valores de X son $D = \{2, 4, 6, 8, \ldots\}$. Como los posibles valores se dieron en secuencia, X es una variable aleatoria discreta. ■

Para estudiar las propiedades básicas de las variables aleatorias discretas, sólo se requieren las herramientas de las matemáticas discretas: sumas y diferencias. El estudio de variables continuas requiere las matemáticas continuas del cálculo: integrales y derivadas.

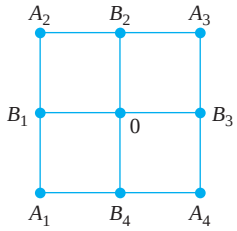
EJERCICIOS Sección 3.1 (1–10)

1. Una viga de concreto puede fallar por esfuerzo cortante (S) o por flexión (F). Suponga que se seleccionan al azar tres vigas que fallaron y se determina el tipo de falla de cada una. Sea $X =$ el número de vigas entre las tres seleccionadas que fallaron por esfuerzo cortante. Ponga en lista cada resultado en el espacio muestral junto con el valor asociado de X.
2. Dé tres ejemplos de variables aleatorias de Bernoulli (aparte de los que aparecen en el texto).
3. Con el experimento del ejemplo 3.3, defina dos variables aleatorias más y mencione los valores posibles de cada una.
4. Sea $X =$ el número de dígitos no cero en un código postal seleccionado al azar. ¿Cuáles son los posibles valores de X? Dé tres posibles resultados y sus valores X asociados.
5. Si el espacio muestral $\mathcal{S}$ es un conjunto infinito, ¿implica esto necesariamente que cualquier variable aleatoria X definida a partir de $\mathcal{S}$ tendrá un conjunto infinito de posibles valores? Si es sí, por qué. Si no, dé un ejemplo.
6. A partir de una hora fija, cada carro que entra a una intersección es observado para ver si da vuelta a la izquierda (I), a la derecha (D) o si sigue de frente (F). El experimento termina en cuanto se observa que un carro da vuelta a la izquierda. Sea $X =$ el número de carros observados. ¿Cuáles son los posibles valores de X? Dé cinco resultados y sus valores X asociados.
7. Para cada variable aleatoria definida aquí, describa el conjunto de posibles valores de la variable y diga si la variable es discreta.

a. X = el número de huevos no quebrados en una caja de huevos estándar seleccionada al azar.
b. Y = el número de estudiantes en una lista de clase de un curso particular que no asisten el primer día de clases.
c. U = el número de veces que un aprendiz tiene que hacerle *swing* a una pelota de golf antes de golpearla.
d. X = la longitud de una serpiente de cascabel seleccionada en forma aleatoria.
e. Z = la cantidad de regalías devengada por la venta de la primera edición de 10,000 libros de texto.
f. Y = el pH de una muestra de suelo elegida al azar.
g. X = la tensión (lb/pulg^2) a la cual una raqueta de tenis seleccionada al azar fue encordada.
h. X = el número total de lanzamientos al aire de una moneda requerido para que tres individuos obtengan una coincidencia (*AAA* o *SSS*).

8. Cada vez que un componente se somete a prueba, ésta es un éxito (S) o una falla (F). Suponga que el componente se prueba repetidamente hasta que ocurre un éxito en tres pruebas *consecutivas*. Sea Y el número necesario de pruebas para lograrlo. Haga una lista de todos los resultados correspondientes a los cinco posibles valores más pequeños de Y y diga qué valor de Y está asociado con cada uno.

9. Un individuo de nombre Claudius se encuentra en el punto 0 del diagrama adjunto.



Con un dispositivo de aleatorización apropiado (tal como un dado tetraédrico, uno que tiene cuatro lados), Claudius primero se mueve a uno de los cuatro lugares B_1, B_2, B_3, B_4. Una vez que está en uno de estos lugares, se utiliza otro dispositivo de aleatorización para decidir si Claudius regresa a 0 o visita uno de los otros dos lugares adyacentes. Este proceso continúa entonces; después de cada movimiento, se determina otro movimiento a uno de los (nuevos) puntos adyacentes lanzando al aire un dado o moneda apropiada.
a. Sea X = el número de movimientos que Claudius hace antes de regresar a 0. ¿Cuáles son los posibles valores de X? ¿Es X discreta o continua?
b. Si también se permiten movimientos a lo largo de los trayectos diagonales que conectan 0 con A_1, A_2, A_3 y A_4, respectivamente, responda las preguntas del inciso (a).

10. Se determinará el número de bombas en uso tanto en la gasolinera de seis bombas como en la gasolinera de cuatro bombas. Dé los posibles valores de cada una de las siguientes variables aleatorias:
a. T = el número total de bombas en uso
b. X = la diferencia entre el número en uso en las gasolineras 1 y 2
c. U = el número máximo de bombas en uso en una u otra gasolinera
d. Z = el número de gasolineras que tienen exactamente dos bombas en uso

3.2 Distribuciones de probabilidad para variables aleatorias discretas

Las probabilidades asignadas a varios resultados en $\mathcal{S}$ determinan a su vez las probabilidades asociadas con los valores de cualquier variable aleatoria X particular. La *distribución de probabilidad de X* dice cómo está distribuida (asignada) la probabilidad total de 1 entre los varios posibles valores de X. Supóngase, por ejemplo, que una empresa acaba de adquirir cuatro impresoras láser y sea X el número de éstas que requieren servicio durante el periodo de garantía. Los posibles valores de X son entonces 0, 1, 2, 3 y 4. La distribución de probabilidad dirá cómo está subdividida la probabilidad de 1 entre estos cinco posibles valores: cuánta probabilidad está asociada con el valor 0 de X, cuánta está adjudicada al valor 1 de X, y así sucesivamente. Se utilizará la siguiente notación para las probabilidades en la distribución:

$$p(0) = \text{la probabilidad del valor 0 de } X = P(X = 0)$$
$$p(1) = \text{la probabilidad del valor 1 de } X = P(X = 1)$$

y así sucesivamente. En general, $p(x)$ denotará la probabilidad asignada al valor de x.

Ejemplo 3.7 El Departamento de Estadística de Cal Poly tiene un laboratorio con seis computadoras reservadas para estudiantes de estadística. Sea X el número de computadoras que están en servicio a una hora particular del día. Suponga que la distribución de probabilidad de X es

como se da en la tabla siguiente; la primera fila de la tabla contiene los posibles valores de X y la segunda da la probabilidad de dicho valor.

x	0	1	2	3	4	5	6
$p(x)$	.05	.10	.15	.25	.20	.15	.10

Ahora se pueden usar propiedades de probabilidad elemental para calcular otras probabilidades de interés. Por ejemplo, la probabilidad de que cuando mucho 2 computadoras estén en servicio es

$$P(X \le 2) = P(X = 0 \text{ o } 1 \text{ o } 2) = p(0) + p(1) + p(2) = .05 + .10 + .15 = .30$$

Como el evento de que *por lo menos 3 computadoras estén en servicio* es complementario a *cuando mucho dos computadoras están en servicio*,

$$P(X \ge 3) = 1 - P(X \le 2) = 1 - .30 = .70$$

la que, desde luego, también se obtiene sumando las probabilidades de los valores 3, 4, 5 y 6. La probabilidad de que entre 2 y 5 computadoras *inclusive* estén en servicio es

$$P(2 \le X \le 5) = P(X = 2, 3, 4 \text{ o } 5) = .15 + .25 + .20 + .15 = .75$$

en tanto que la probabilidad de que el número de computadoras en servicio esté *estrictamente entre* 2 y 5 es

$$P(2 < X < 5) = P(X = 3 \text{ o } 4) = .25 + .20 = .45$$ ■

DEFINICIÓN

La **distribución de probabilidad** o **función de masa de probabilidad** (fmp) de una variable discreta se define para cada número x como $p(x) = P(X = x) = P$ (todas las $s \in \mathcal{S}$: $X(s) = x$).

En palabras, para cada valor posible x de la variable aleatoria, la función de masa de probabilidad especifica la probabilidad de observar dicho valor cuando se realiza el experimento. Se requieren las condiciones $p(x) \ge 0$ y $\sum_{\text{todas las } x \text{ posibles}} p(x) = 1$ de cualquier función de masa de probabilidad.

La función de masa de probabilidad de X en el ejemplo previo se dio simplemente en la descripción del problema. A continuación se consideran varios ejemplos en los cuales se explotan varias propiedades de probabilidad para obtener la distribución deseada.

Ejemplo 3.8 Seis lotes de componentes están listos para ser enviados por un proveedor. El número de componentes defectuosos en cada lote es como sigue:

Lote	1	2	3	4	5	6
Número de componentes defectuosos	0	2	0	1	2	0

Uno de estos lotes tiene que ser seleccionado al azar para ser enviado a un cliente particular. Sea X el número de defectuosos en el lote seleccionado. Los tres posibles valores de X son 0, 1 y 2. De los seis eventos simples igualmente probables, tres dan por resultado $X = 0$, uno $X = 1$ y los otros dos $X = 2$. Entonces

$$p(0) = P(X = 0) = P(\text{el lote 1 o 3 o 6 es enviado}) = \frac{3}{6} = .500$$

$$p(1) = P(X = 1) = P(\text{el lote 4 es enviado}) = \frac{1}{6} = .167$$

$$p(2) = P(X = 2) = P(\text{el lote 2 o 5 es enviado}) = \frac{2}{6} = .333$$

Es decir, una probabilidad de .500 se asigna al valor 0 de X, una probabilidad de .167 se asigna al valor 1 de X, y la probabilidad restante, .333, se asocia con el valor 2 de X. Los valores de X junto con sus probabilidades especifican la función de masa de probabilidad. Si este experimento se repitiera una y otra vez, a la larga $X = 0$ ocurriría la mitad del tiempo, $X = 1$ un sexto del tiempo y $X = 2$ un tercio del tiempo. ■

Ejemplo 3.9 Considere si la siguiente persona que compre una computadora en cierta tienda de electrónicos elegirá un modelo portátil o uno de escritorio. Sea

$$X = \begin{cases} 1 & \text{si el cliente compra una computadora de escritorio} \\ 0 & \text{si el cliente compra una computadora portátil} \end{cases}$$

Si 20% de todos los compradores durante esa semana seleccionan una de escritorio, la función de masa de probabilidad de X es

$$p(0) = P(X = 0) = P(\text{el siguiente cliente compra un modelo portátil}) = .8$$
$$p(1) = P(X = 1) = P(\text{el siguiente cliente compra un modelo de escritorio}) = .2$$
$$p(x) = P(X = x) = 0 \text{ con } x \neq 0 \text{ o } 1$$

Una descripción equivalente es

$$p(x) = \begin{cases} .8 & \text{si } x = 0 \\ .2 & \text{si } x = 1 \\ 0 & \text{si } x \neq 0 \text{ o } 1 \end{cases}$$

La figura 3.2 es una ilustración de esta función de masa de probabilidad, llamada *gráfica lineal*. X es, desde luego, una variable aleatoria de Bernoulli y $p(x)$ es una función de masa de probabilidad de Bernoulli.

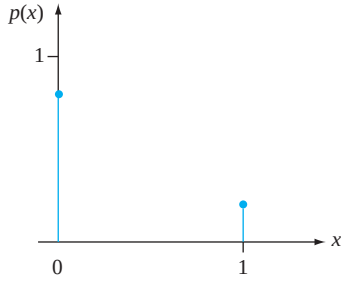


Figura 3.2 Gráfica lineal para la función de masa de probabilidad de Bernoulli en el ejemplo 3.9 ■

Ejemplo 3.10 Considere un grupo de cinco donadores de sangre potenciales, a, b, c, d y e, de los cuales sólo a y b tienen sangre tipo O+. Se determinará en orden aleatorio el tipo de sangre con cinco muestras, una de cada individuo, hasta que se identifique un individuo O+. Sea la variable aleatoria $Y =$ el número de exámenes de sangre para identificar un individuo O+. Entonces la función de masa de probabilidad de Y es

$$p(1) = P(Y = 1) = P(a \text{ o } b \text{ examinados primero}) = \frac{2}{5} = .4$$

$$p(2) = P(Y = 2) = P(c, d \text{ o } e \text{ primero, y luego } a \text{ o } b)$$
$$= P(c, d, \text{ o } e \text{ primero}) \cdot P(a \text{ o } b \text{ a continuación} \mid c, d \text{ o } e \text{ primero}) = \frac{3}{5} \cdot \frac{2}{4} = .3$$

$$p(3) = P(Y = 3) = P(c, d \text{ o } e \text{ primero y segundo, y luego } a \text{ o } b)$$
$$= \left(\frac{3}{5}\right)\left(\frac{2}{4}\right)\left(\frac{2}{3}\right) = .2$$

$$p(4) = P(Y = 4) = P(c, d \text{ y } e \text{ primero}) = \left(\frac{3}{5}\right)\left(\frac{2}{4}\right)\left(\frac{1}{3}\right) = .1$$

$$p(y) = 0 \text{ si } y \neq 1, 2, 3, 4$$

En forma tabular, la función de masa de probabilidad es

y	1	2	3	4
$p(y)$	.4	.3	.2	.1

donde cualquier valor de y que no aparece en la tabla recibe cero probabilidad. La figura 3.3 muestra una gráfica lineal de la función de masa de probabilidad.

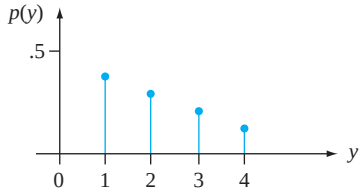


Figura 3.3 Gráfica lineal para la función de masa de probabilidad de Bernoulli del ejemplo 3.10 ■

Un modelo utilizado en física para un sistema de "masas puntuales" sugirió el nombre "función de masa de probabilidad". En este modelo, las masas están distribuidas en varios lugares x a lo largo de un eje unidimensional. La función de masa de probabilidad describe cómo está distribuida la masa de probabilidad total de 1 en varios puntos a lo largo del eje de posibles valores de la variable aleatoria (dónde y cuánta masa hay en cada x).

Otra representación pictórica útil de una función de masa de probabilidad, llamada **histograma de probabilidad,** es similar a los histogramas discutidos en el capítulo 1. Sobre cada y con $p(y) > 0$, se construye un rectángulo con su centro en y. La altura de cada rectángulo es proporcional a $p(y)$ y la base es la misma para todos los rectángulos. Cuando los valores posibles están equidistantes, con frecuencia se selecciona la base como la distancia entre valores y sucesivos (aunque podría ser más pequeña). La figura 3.4 muestra dos histogramas de probabilidad.

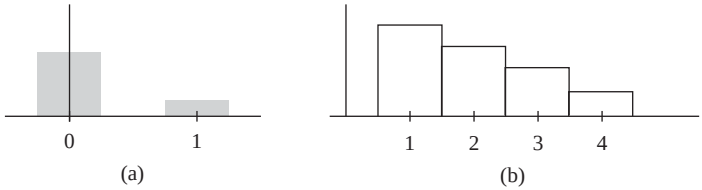


Figura 3.4 Histogramas de probabilidad: (a) Ejemplo 3.9; (b) Ejemplo 3.10

A menudo es útil pensar en una función de masa de probabilidad como un modelo matemático de una población discreta.

Ejemplo 3.11 Considere seleccionar al azar un estudiante de entre los 15,000 inscritos en el semestre actual en la Universidad Mega. Sea X = el número de cursos en los cuales el estudiante seleccionado está inscrito y suponga que X tiene la siguiente función de masa de probabilidad:

x	1	2	3	4	5	6	7
$p(x)$	.01	.03	.13	.25	.39	.17	.02

Una forma de ver esta situación es pensar en la población como compuesta de 15,000 individuos, cada uno con su propio valor X; la proporción con cada valor de X está dada por $p(x)$. Un punto de vista alternativo es olvidarse de los estudiantes y pensar en la población como compuesta de los valores X: existen algunos 1 en la población, algunos 2, . . . , y finalmente algunos 7. La población se compone entonces de los números 1, 2, . . . , 7 (por lo tanto es discreta) y $p(x)$ da un modelo para la distribución de los valores de población. ■

Una vez que se tiene el modelo de la población, se utilizará para calcular valores de características de la población (p. ej., la media μ) y para hacer inferencias sobre tales características.

Parámetro de una distribución de probabilidad

La función de masa de probabilidad de Bernoulli en el ejemplo 3.9, fue $p(0) = .8$ y $p(1) = .2$ porque 20% de todos los compradores seleccionaron una computadora de escritorio. En otro almacén, puede ser el caso que $p(0) = .9$ y $p(1) = .1$. Más generalmente, la función de masa de probabilidad de cualquier variable aleatoria de Bernoulli puede ser expresada en la forma $p(1) = \alpha$ y $p(0) = 1 - \alpha$, donde $0 < \alpha < 1$. Como la función de masa de probabilidad depende del valor particular de α, con frecuencia se escribe $p(x; \alpha)$ en lugar de sólo $p(x)$:

$$p(x;\alpha) = \begin{cases} 1-\alpha & \text{si } x = 0 \\ \alpha & \text{si } x = 1 \\ 0 & \text{de lo contrario} \end{cases} \tag{3.1}$$

Entonces cada opción de α en la expresión (3.1) da una función de masa de probabilidad diferente.

DEFINICIÓN

Supóngase que $p(x)$ depende de la cantidad que puede ser asignada a cualquiera de un número de valores posibles, y cada valor determina una distribución de probabilidad diferente. Tal cantidad se llama **parámetro** de distribución. El conjunto de todas las distribuciones de probabilidad para diferentes valores del parámetro se llama **familia** de distribuciones de probabilidad.

La cantidad α en la expresión (3.1) es un parámetro. Cada número diferente α entre 0 y 1 determina un miembro diferente de la familia de distribuciones de Bernoulli.

Ejemplo 3.12 A partir de un tiempo fijo, se observa el sexo de cada niño recién nacido en un hospital hasta que nace un varón (B). Sea $p = P(B)$ y suponga que los nacimientos sucesivos son independientes y defina la variable aleatoria X como $x =$ número de nacimientos observados. Entonces

$$p(1) = P(X = 1) = P(B) = p$$
$$p(2) = P(X = 2) = P(GB) = P(G) \cdot P(B) = (1 - p)p$$

y

$$p(3) = P(X = 3) = P(GGB) = P(G) \cdot P(G) \cdot P(B) = (1 - p)^2 p$$

Continuando de esta manera, emerge una fórmula general:

$$p(x) = \begin{cases} (1-p)^{x-1}p & x = 1, 2, 3, \ldots \\ 0 & \text{de lo contrario} \end{cases} \tag{3.2}$$

El parámetro p puede asumir cualquier valor entre 0 y 1. La expresión (3.2) describe la familia de distribuciones *geométricas*. En el ejemplo del sexo, $p = .51$ podría ser apropiado, pero si estábamos buscando el primer hijo con sangre Rh positiva, entonces podríamos tener $p = .85$. ■

Función de distribución acumulativa

Para algún valor fijo x, a menudo se desea calcular la probabilidad de que el valor observado de X será cuando mucho x. Por ejemplo, la función de masa de probabilidad en el ejemplo 3.8 fue

$$p(x) = \begin{cases} .500 & x = 0 \\ .167 & x = 1 \\ .333 & x = 2 \\ 0 & \text{de lo contrario} \end{cases}$$

La probabilidad de que X sea cuando mucho 1 es entonces

$$P(X \le 1) = p(0) + p(1) = .500 + .167 = .667$$

En este ejemplo, $X \le 1.5$ si y sólo si $X \le 1$, por lo tanto

$$P(X \le 1.5) = P(X \le 1) = .667$$

Asimismo,

$$P(X \le 0) = P(X = 0) = .5, \quad P(X \le .75) = .5$$

Y de hecho con cualquier x que satisfaga $0 \le x < 1$, $P(X \le x) = .5$. El valor de X más grande posible es 2, por lo tanto

$$P(X \le 2) = 1, \quad P(X \le 3.7) = 1, \quad P(X \le 20.5) = 1$$

y así sucesivamente. Obsérvese que $P(X < 1) < P(X \le 1)$ puesto que la segunda parte de la desigualdad incluye la probabilidad del valor 1 de X, en tanto que la primera no. Más generalmente, cuando X es discreta y x es un valor posible de la variable, $P(X < x) < P(X \le x)$.

DEFINICIÓN

La **función de distribución acumulativa** (fda) $F(x)$ de una variable aleatoria discreta X con función de masa de probabilidad $p(x)$ se define para cada número x como

$$F(x) = P(X \le x) = \sum_{y:y \le x} p(y) \tag{3.3}$$

Para cualquier número x, $F(x)$ es la probabilidad de que el valor observado de X será cuando mucho x.

Ejemplo 3.13 Una tienda vende unidades de memoria flash, ya sea con 1 GB, 2 GB, 4 GB, 8 GB o 16 GB de memoria. La tabla adjunta muestra la distribución de Y = la cantidad de memoria en un disco comprado:

y	1	2	4	8	16
$p(y)$	.05	.10	.35	.40	.10

Primero se determina $F(y)$ para cada uno de los cinco valores posibles de Y:

$$\begin{aligned} F(1) &= P(Y \le 1) = P(Y = 1) = p(1) = .05 \\ F(2) &= P(Y \le 2) = P(Y = 1 \text{ o } 2) = p(1) + p(2) = .15 \\ F(4) &= P(Y \le 4) = P(Y = 1 \text{ o } 2 \text{ o } 4) = p(1) + p(2) + p(4) = .50 \\ F(8) &= P(Y \le 8) = p(1) + p(2) + p(4) + p(8) = .90 \\ F(16) &= P(Y \le 16) = 1 \end{aligned}$$

Ahora con cualquier otro número y, $F(y)$ será igual al valor de F en el valor más próximo posible de Y a la izquierda de y. Por ejemplo,

$$\begin{aligned} F(2.7) &= P(Y \le 2.7) = P(Y \le 2) = F(2) = .15 \\ F(7.999) &= P(Y \le 7.999) = P(Y \le 4) = F(4) = .50 \end{aligned}$$

Si y es menor que 1, $F(y) = 0$ [por ejemplo, $F(.58) = 0$], y si y es por lo menos 16, $F(y) = 1$ [por ejemplo, $F(25) = 1$]. La fda es, pues,

$$F(y) = \begin{cases} 0 & y < 1 \\ .05 & 1 \le y < 2 \\ .15 & 2 \le y < 4 \\ .50 & 4 \le y < 8 \\ .90 & 8 \le y < 16 \\ 1 & 16 \le y \end{cases}$$

En la figura 3.5 se muestra una gráfica de esta fda.

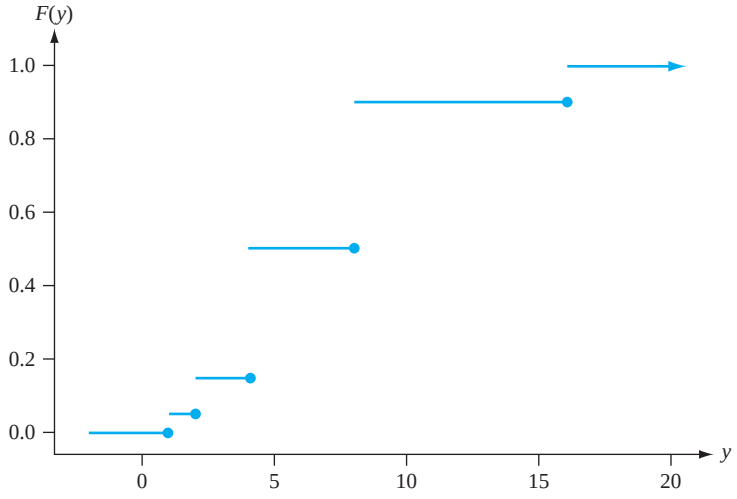


Figura 3.5 Gráfica de la función de distribución acumulativa del ejemplo 3.13 ■

Para una variable aleatoria discreta X, la gráfica de $F(x)$ mostrará un salto con cada valor posible de X, y será plana entre los valores posibles. Tal gráfica se conoce como **función escalón.**

Ejemplo 3.14 (Continuación del ejemplo 3.12)

La fmp de $X =$ *el número de nacimientos* tenía la forma

$$p(x) = \begin{cases} (1-p)^{x-1}p & x = 1, 2, 3, \ldots \\ 0 & \text{de lo contrario} \end{cases}$$

Para cualquier entero positivo x,

$$F(x) = \sum_{y \le x} p(y) = \sum_{y=1}^{x} (1-p)^{y-1}p = p\sum_{y=0}^{x-1}(1-p)^{y} \qquad (3.4)$$

Para evaluar esta suma, se utiliza el hecho de que la suma parcial de una serie geométrica es

$$\sum_{y=0}^{k} a^y = \frac{1-a^{k+1}}{1-a}$$

Utilizando esto en la ecuación (3.4), con $a = 1 - p$ y $k = x - 1$, se obtiene

$$F(x) = p \cdot \frac{1 - (1 - p)^x}{1 - (1 - p)} = 1 - (1 - p)^x \quad x \text{ es un entero positivo}$$

Como F es una constante entre enteros positivos,

$$F(x) = \begin{cases} 0 & x < 1 \\ 1 - (1 - p)^{[x]} & x \geq 1 \end{cases} \tag{3.5}$$

donde $[x]$ es el entero más grande $\leq x$ (p. ej., $[2.7] = 2$). Así pues, si $p = .51$ como en el ejemplo de los nacimientos, entonces la probabilidad de tener que examinar cuando mucho cinco nacimientos para ver el primer varón es $F(5) = 1 - (.49)^5 = 1 - .0282 = .9718$, mientras que $F(10) \approx 1.0000$. Esta función de distribución acumulativa se ilustra en la figura 3.6.

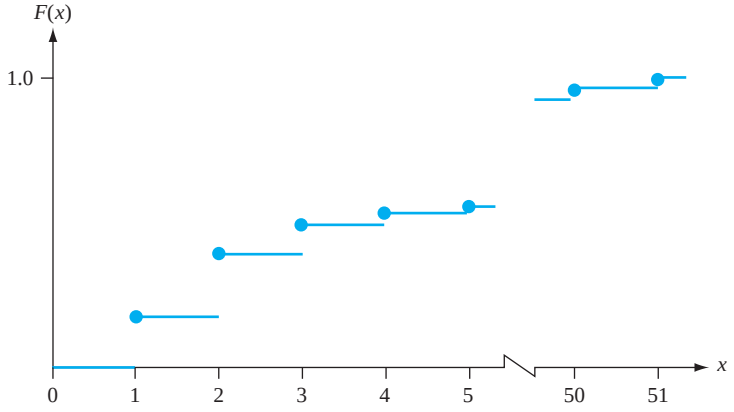


Figura 3.6 Gráfica de $F(x)$ para el ejemplo 3.14 ■

En los ejemplos presentados hasta ahora, la función de distribución acumulativa se dedujo de la función de masa de probabilidad. Este proceso puede ser invertido para obtener la función de masa de probabilidad a partir de la función de distribución acumulativa siempre que ésta esté disponible. Por ejemplo, considérese otra vez la variable aleatoria del ejemplo 3.7 (el número de computadoras usadas en un laboratorio); los valores posibles de X son 0, 1, . . . , 6. Entonces

$$\begin{aligned} p(3) &= P(X = 3) \\ &= [p(0) + p(1) + p(2) + p(3)] - [p(0) + p(1) + p(2)] \\ &= P(X \leq 3) - P(X \leq 2) \\ &= F(3) - F(2) \end{aligned}$$

Más generalmente, la probabilidad de que X quede dentro de un intervalo especificado es fácil de obtener a partir de la función de distribución acumulativa. Por ejemplo,

$$\begin{aligned} P(2 \leq X \leq 4) &= p(2) + p(3) + p(4) \\ &= [p(0) + \cdots + p(4)] - [p(0) + p(1)] \\ &= P(X \leq 4) - P(X \leq 1) \\ &= F(4) - F(1) \end{aligned}$$

Obsérvese que $P(2 \leq X \leq 4) \neq F(4) - F(2)$. Esto es porque el valor 2 de X está incluido en $2 \leq X \leq 4$, así que no se desea restar su probabilidad. Sin embargo, $P(2 < X \leq 4) = F(4) - F(2)$ porque $X = 2$ no está incluido en el intervalo $2 < X \leq 4$.

PROPOSICIÓN

Para dos números cualesquiera a y b con $a \le b$,

$$P(a \le X \le b) = F(b) - F(a-)$$

donde "$a-$" representa el valor posible de X más grande que es estrictamente menor que a. En particular, si los únicos valores posibles son enteros, y si a y b son enteros, entonces

$$\begin{aligned} P(a \le X \le b) &= P(X = a \text{ o } a + 1 \text{ o} \ldots \text{ o } b) \\ &= F(b) - F(a - 1) \end{aligned}$$

Con $a = b$ se obtiene $P(X = a) = F(a) - F(a - 1)$ en este caso.

La razón de restar $F(a-)$ en lugar de $F(a)$ es que se desea incluir $P(X = a)$; $F(b) - F(a)$ da $P(a < X \le b)$. Esta proposición se utilizará extensamente cuando se calculen las probabilidades binomial y de Poisson en las secciones 3.4 y 3.6.

Ejemplo 3.15 Sea X = el número de días de ausencia por enfermedad tomados por un empleado seleccionado al azar de una gran compañía durante un año particular. Si el número máximo de días de ausencia por enfermedad permisibles al año es de 14, los valores posibles de X son 0, 1, . . . , 14. Con $F(0) = .58$, $F(1) = .72$, $F(2) = .76$, $F(3) = .81$, $F(4) = .88$ y $F(5) = .94$,

$$P(2 \le X \le 5) = P(X = 2, 3, 4 \text{ o } 5) = F(5) - F(1) = .22$$

y

$$P(X = 3) = F(3) - F(2) = .05$$ ■

EJERCICIOS Sección 3.2 (11–28)

11. En un taller de servicio automotriz especializado en afinaciones se sabe que 45% de todas las afinaciones se realizan en automóviles de cuatro cilindros, 40% en automóviles de seis cilindros y 15% en automóviles de ocho cilindros. Sea X = el número de cilindros en el siguiente carro que va a ser afinado.
- **a.** ¿Cuál es la función de masa de probabilidad de X?
- **b.** Trace una gráfica lineal y un histograma de probabilidad de la función de masa de probabilidad del inciso (a).
- **c.** ¿Cuál es la probabilidad de que el siguiente carro afinado sea de por lo menos seis cilindros? ¿Más de seis cilindros?

12. Las líneas aéreas en ocasiones venden boletos de más. Suponga que para un avión de 50 asientos, 55 pasajeros tienen boleto. Defina la variable aleatoria Y como el número de pasajeros con boleto que en realidad se presentan para el vuelo. La función de masa de probabilidad de Y aparece en la tabla adjunta.

y	45	46	47	48	49	50	51	52	53	54	55
$p(y)$	.05	.10	.12	.14	.25	.17	.06	.05	.03	.02	.01

- **a.** ¿Cuál es la probabilidad de que el vuelo acomode a todos los pasajeros con boleto que se presenten?
- **b.** ¿Cuál es la probabilidad de que no todos los pasajeros con boleto que aparecieron puedan ser acomodados?
- **c.** Si usted es la primera persona en la lista de espera (lo que significa que será el primero en abordar el avión si hay boletos disponibles después de que todos los pasajeros con boleto hayan sido acomodados), ¿cuál es la probabilidad de que pueda tomar el vuelo? ¿Cuál es esta probabilidad si usted es la tercera persona en la lista de espera?

13. Una empresa de ventas en línea dispone de seis líneas telefónicas. Sea X el número de líneas en uso en un tiempo especificado. Suponga que la función de masa de probabilidad de X es la que se da en la tabla adjunta.

x	0	1	2	3	4	5	6
$p(x)$	.10	.15	.20	.25	.20	.06	.04

Calcule la probabilidad de cada uno de los siguientes eventos.
- **a.** {cuando mucho tres líneas están en uso}
- **b.** {menos de tres líneas están en uso}
- **c.** {por lo menos tres líneas están en uso}
- **d.** {entre dos y cinco líneas, inclusive, están en uso}
- **e.** {entre dos y cuatro líneas, inclusive, no están en uso}
- **f.** {por lo menos cuatro líneas no están en uso}

14. El departamento de planeación de un condado requiere que un contratista presente uno, dos, tres, cuatro o cinco formas (según la naturaleza del proyecto) para solicitar un permiso de construcción. Sea Y = número de formas requeridas del siguiente solicitante. Se sabe que la probabilidad de que se requieran y formas es proporcional a y, es decir, $p(y) = ky$ con $y = 1, \ldots, 5$.

- **a.** ¿Cuál es el valor de k? [*Sugerencia*: $\sum_{y=1}^{5} p(y) = 1$.]
- **b.** ¿Cuál es la probabilidad de que cuando mucho se requieran tres formas?
- **c.** ¿Cuál es la probabilidad de que se requieran entre dos y cuatro formas (inclusive)?
- **d.** ¿Podría ser $p(y) = y^2/50$ con $y = 1, \ldots, 5$ la función de masa de probabilidad de Y?

15. Muchos fabricantes cuentan con programas de control de calidad que incluyen la inspección de los materiales recibidos en busca de defectos. Suponga que un fabricante de computadoras recibe tarjetas madre en lotes de cinco. Se seleccionan dos tarjetas de cada lote para inspeccionarlas. Se puede representar los posibles resultados del proceso de selección por pares. Por ejemplo, el par (1, 2) representa la selección de las tarjetas 1 y 2 para inspección.

- **a.** Mencione los diez posibles resultados diferentes.
- **b.** Suponga que las tarjetas 1 y 2 son las únicas defectuosas en un lote de cinco. Dos tarjetas tienen que ser seleccionadas al azar. Defina X como el número de tarjetas defectuosas observadas entre las inspeccionadas. Encuentre la distribución de probabilidad de X.
- **c.** Sea $F(x)$ la función de distribución acumulativa de X. Primero determine $F(0) = P(X \leq 0)$, $F(1)$ y $F(2)$; luego obtenga $F(x)$ para todas las demás x.

16. Algunas partes de California son particularmente propensas a los temblores. Suponga que en un área metropolitana, 25% de todos los propietarios de casa están asegurados contra daños provocados por terremotos. Se seleccionan al azar cuatro propietarios de casa; sea X el número entre los cuatro que están asegurados contra terremotos.

- **a.** Encuentre la distribución de probabilidad de X. [*Sugerencia*: Sea S un propietario de casa asegurado y F uno no asegurado. Entonces un posible resultado es $SFSS$, con probabilidad (.25)(.75)(.25)(.25) y el valor 3 de X asociado. Existen otros 15 resultados.]
- **b.** Trace el histograma de probabilidad correspondiente.
- **c.** ¿Cuál es el valor más probable de X?
- **d.** ¿Cuál es la probabilidad de que por lo menos dos de los cuatro seleccionados estén asegurados contra terremotos?

17. El voltaje de una batería nueva puede ser aceptable (A) o inaceptable (I). Una linterna requiere dos baterías, así que las baterías serán independientemente seleccionadas y probadas hasta encontrar dos aceptables. Suponga que 90% de todas las baterías tienen voltajes aceptables. Sea Y el número de baterías que deben ser probadas.

- **a.** ¿Cuál es $p(2)$, es decir, $P(Y = 2)$?
- **b.** ¿Cuál es $p(3)$? [*Sugerencia*: existen dos resultados diferentes que producen $Y = 3$.]
- **c.** Para tener $Y = 5$, ¿qué debe ser cierto de la quinta batería seleccionada? Mencione los cuatro resultados con los cuales $Y = 5$ y luego determine $p(5)$.
- **d.** Use el patrón de sus respuestas en los incisos (a)–(c) para obtener una fórmula general para $p(y)$.

18. Dos dados de seis caras son lanzados al aire en forma independiente. Sea M = el máximo de los dos lanzamientos (por lo tanto $M(1, 5) = 5$, $M(3, 3) = 3$, etcétera).

- **a.** ¿Cuál es la función de masa de probabilidad de M? [*Sugerencia*: primero determine $p(1)$, luego $p(2)$, y así sucesivamente.]
- **b.** Determine la función de distribución acumulativa de M y grafíquela.

19. Una biblioteca se suscribe a dos revistas diferentes de noticias semanales, cada una de las cuales se supone que llega en el correo de los miércoles. En realidad, cada una puede llegar el miércoles, jueves, viernes o sábado. Suponga que las dos llegan independientemente una de otra y para cada una P(mié) = .3, P(jue) = .4, P(vie) = .2 y P(sáb) = .1. Sea Y = el número de días después del miércoles que pasan para que ambas revistas lleguen (por lo tanto los posibles valores de Y son 0, 1, 2, o 3). Calcule la función de masa de probabilidad de Y. [*Sugerencia*: hay 16 posibles resultados: $Y(M, M) = 0$, $Y(V, J) = 2$, y así sucesivamente.]

20. Tres parejas y dos individuos solteros han sido invitados a un seminario de inversión y han aceptado asistir. Suponga que la probabilidad de que cualquier pareja o individuo particular llegue tarde es de .4 (una pareja viajará en el mismo vehículo, así que ambos llegarán a tiempo o bien ambos llegarán tarde). Suponga que diferentes parejas e individuos llegan puntuales o tarde independientemente unos de otros. Sea X = el número de personas que llegan tarde al seminario.

- **a.** Determine la función de masa de probabilidad de X. [*Sugerencia*: designe las tres parejas #1, #2 y #3 y los dos individuos #4 y #5.]
- **b.** Obtenga la función de distribución acumulativa de X y úsela para calcular $P(2 \leq X \leq 6)$.

21. Suponga que lee los números de este año del *New York Times* y que anota cada número que aparece en un artículo de noticias: el ingreso de un oficial ejecutivo en jefe, el número de cajas de vino producidas por una compañía vinícola, la contribución caritativa total de un político durante el año fiscal previo, la edad de una celebridad, y así sucesivamente. Ahora enfóquese en el primer dígito de cada número, el cual podría ser 1, 2, . . . , 8 o 9. Su primer pensamiento podría que el primer dígito X de un número seleccionado al azar sería igualmente probable que fuera una de las nueve posibilidades (una distribución uniforme discreta). Sin embargo, mucha evidencia empírica así como también algunos argumentos teóricos sugieren una distribución de probabilidad alternativa llamada *ley de Benford*:

$$p(x) = P(\text{el primer dígito es } x) = \log_{10}\left(\frac{x+1}{x}\right) \quad x = 1, 2, \ldots, 9$$

- **a.** Sin calcular probabilidades individuales de esta fórmula, demuestre que especifica una función de masa de probabilidad legítima.
- **b.** Ahora calcule las probabilidades individuales y compare con la distribución uniforme discreta correspondiente.
- **c.** Obtenga la función de distribución acumulativa de X.
- **d.** Utilizando la función de distribución acumulativa, ¿cuál es la probabilidad de que el primer dígito sea cuando mucho 3? ¿Por lo menos 5?

[*Nota*: la ley de Benford es la base de algunos procedimientos de auditoría utilizados para detectar fraudes en reportes financieros, por ejemplo, por el Servicio de Ingresos Internos.]

22. Remítase al ejercicio 13 y calcule y trace la gráfica de la función de distribución acumulativa $F(x)$. Luego utilícela para calcular las probabilidades de los eventos dados en los incisos (a)–(d) de dicho problema.

23. Una organización de protección al consumidor que habitualmente evalúa automóviles nuevos reporta el número de defectos importantes encontrados en cada carro examinado. Sea X el número de defectos importantes en un carro seleccionado al azar de cierto tipo. La función de distribución acumulativa de X es la siguiente:

$$F(x) = \begin{cases} 0 & x < 0 \\ .06 & 0 \le x < 1 \\ .19 & 1 \le x < 2 \\ .39 & 2 \le x < 3 \\ .67 & 3 \le x < 4 \\ .92 & 4 \le x < 5 \\ .97 & 5 \le x < 6 \\ 1 & 6 \le x \end{cases}$$

Calcule las siguientes probabilidades directamente con la función de distribución acumulativa:

a. $p(2)$, es decir, $P(X = 2)$ **b.** $P(X > 3)$
c. $P(2 \le X \le 5)$ **d.** $P(2 < X < 5)$

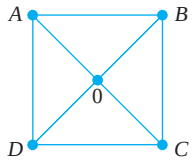
24. Una compañía de seguros ofrece a sus asegurados varias opciones diferentes de pago de primas. Para un asegurado seleccionado al azar, sea $X =$ el número de meses entre pagos sucesivos. La función de distribución acumulativa es la siguiente:

$$F(x) = \begin{cases} 0 & x < 1 \\ .30 & 1 \le x < 3 \\ .40 & 3 \le x < 4 \\ .45 & 4 \le x < 6 \\ .60 & 6 \le x < 12 \\ 1 & 12 \le x \end{cases}$$

a. ¿Cuál es la función de masa de probabilidad de X?
b. Con sólo la función de distribución acumulativa, calcule $P(3 \le X \le 6)$ y $P(4 \le X)$.

25. En el ejemplo 3.12, sea $Y =$ el número de niñas nacidas antes de que termine el experimento. Con $p = P(B)$ y $1 - p = P(G)$, ¿cuál es la función de masa de probabilidad de Y? [*Sugerencia*: primero ponga en lista los posibles valores de Y, inicie con el más pequeño y continúe hasta que encuentre una fórmula general.]

26. Alvie Singer vive en 0 en el diagrama adjunto y sus cuatro amigos viven en A, B, C y D. Un día Alvie decide visitarlos, así que lanza al aire una moneda imparcial dos veces para decidir a cuál de los cuatro visitar. Una vez que está en la casa de uno de sus amigos, regresará a su casa o bien proseguirá a una de las dos casas adyacentes (tales como 0, A o C, cuando está en B) con cada una de las tres posibilidades cuya probabilidad es $\frac{1}{3}$. De este modo, Alvie continúa visitando a sus amigos hasta que regresa a casa.



a. Sea $X =$ el número de veces que Alvie visita a un amigo. Obtenga la función de masa de probabilidad de X.
b. Sea $Y =$ el número de segmentos de línea recta que Alvie recorre (incluidos los que conducen a 0 o que parten de ahí). ¿Cuál es la función de masa de probabilidad de Y?
c. Suponga que sus amigas viven en A y C y sus amigos en B y D. Si $Z =$ el número de visitas a amigas, ¿cuál es la función de masa de probabilidad de Z?

27. Después de que todos los estudiantes salieron del salón de clases, un profesor de estadística nota que cuatro ejemplares del texto se quedaron debajo de los escritorios. Al principio de la siguiente clase, el profesor distribuye los cuatro libros al azar a cada uno de los cuatro estudiantes (1, 2, 3 y 4) que dicen haber dejado los libros. Un posible resultado es que 1 reciba el libro de 2, que 2 reciba el libro de 4 y que 3 reciba su propio libro y que 4 reciba el libro de 1. Este resultado puede ser abreviado como (2, 4, 3, 1).

a. Mencione los otros 23 resultados posibles.
b. Si X es el número de estudiantes que reciben su propio libro, determine la función de masa de probabilidad de X.

28. Demuestre que la función de distribución acumulativa de $F(x)$ es no decreciente; es decir, $x_1 < x_2$ implica que $F(x_1) \le F(x_2)$. ¿En qué condición será $F(x_1) = F(x_2)$?

3.3 Valores esperados

Considérese una universidad que tiene 15,000 estudiantes y sea $X =$ el número de cursos en los cuales está inscrito un estudiante seleccionado al azar. La función de masa de probabilidad de X se determina como sigue. Como $p(1) = .01$, se sabe que $(.01) \cdot (15{,}000) = 150$ de los estudiantes están inscritos en un curso y asimismo con los demás valores de x.

x	1	2	3	4	5	6	7
$p(x)$	.01	.03	.13	.25	.39	.17	.02
Número de inscritos	150	450	1950	3750	5850	2550	300

(3.6)

El número promedio de cursos por estudiante o el valor promedio de X en la población se obtiene al calcular el número total de cursos tomados por todos los estudiantes y dividir entre el número total de estudiantes. Como cada uno de los 150 estudiantes está tomando un curso, estos 150 contribuyen con 150 cursos al total. Asimismo, 450 estudiantes contribuyen con 2(450) cursos, y así sucesivamente. El valor promedio de la población de X es entonces

$$\frac{1(150) + 2(450) + 3(1950) + \cdots + 7(300)}{15{,}000} = 4.57 \tag{3.7}$$

Como 150/15,000 = .01 = $p(1)$, 450/15,000 = .03 = $p(2)$, y así sucesivamente, una expresión alternativa para (3.7) es

$$1 \cdot p(1) + 2 \cdot p(2) + \cdots + 7 \cdot p(7) \tag{3.8}$$

La expresión (3.8) muestra que para calcular el valor promedio de la población de X, sólo se necesitan los valores posibles de X junto con sus probabilidades (proporciones). En particular, el tamaño de la población no viene al caso en tanto la función de masa de probabilidad esté dada por (3.6). El valor promedio o medio de X es entonces el promedio *ponderado* de los posibles valores 1, . . . , 7, donde las ponderaciones son las probabilidades de esos valores.

Valor esperado de X

DEFINICIÓN

Sea X una variable aleatoria discreta con un conjunto de valores posibles D y una función de masa de probabilidad $p(x)$. El **valor esperado** o **valor medio** de X, denotado por $E(X)$ o μ_X o sólo μ, es

$$E(X) = \mu_X = \sum_{x \in D} x \cdot p(x)$$

Ejemplo 3.16 Para la función de masa de probabilidad de X = *número de cursos* en (3.6),

$$\begin{aligned}\mu &= 1 \cdot p(1) + 2 \cdot p(2) + \cdots + 7 \cdot p(7)\\ &= (1)(.01) + 2(.03) + \cdots + (7)(.02)\\ &= .01 + .06 + .39 + 1.00 + 1.95 + 1.02 + .14 = 4.57\end{aligned}$$

Si se piensa en la población como compuesta de los valores 1, 2, . . . , 7 de X, entonces $\mu = 4.57$ es la media de la población. En lo que sigue, a menudo se hará referencia a μ como la *media de la población* en lugar de la media de X en la población. Tenga en cuenta que μ aquí no es 4, el promedio normal de 1, . . . , 7, porque la distribución pone más peso en 4, 5 y 6 que en otros valores de X. ■

En el ejemplo 3.16, el valor esperado μ fue 4.57, el cual no es un valor posible de X. La palabra *esperado* deberá interpretarse con precaución porque no se esperaría ver un valor de X de 4.57 cuando se selecciona un solo estudiante.

Ejemplo 3.17 Exactamente después de nacer, cada recién nacido es evaluado en una escala llamada escala de Apgar. Las evaluaciones posibles son 0, 1, . . . , 10, con la evaluación del niño determinada por color, tono muscular, esfuerzo para respirar, ritmo cardiaco e irritabilidad refleja (la mejor evaluación posible es 10). Sea X la evaluación Apgar de un niño seleccionado al azar nacido en cierto hospital durante el siguiente año y supóngase que la función de masa de probabilidad de X es

x	0	1	2	3	4	5	6	7	8	9	10
$p(x)$	.002	.001	.002	.005	.02	.04	.18	.37	.25	.12	.01

Entonces el valor medio de X es

$$\begin{aligned} E(X) = \mu &= 0(.002) + 1(.001) + 2(.002) \\ &\quad + \cdots + 8(.25) + 9(.12) + 10(.01) \\ &= 7.15 \end{aligned}$$

De nuevo, μ no es un valor posible de la variable X. Además, como la variable se refiere a un niño futuro, no existe ninguna población concreta a la cual se podría referir μ. En cambio, la función de masa de probabilidad se considera como un modelo de una población conceptual compuesta de los valores 0, 1, 2, . . . , 10. El valor medio de esta población conceptual es entonces $\mu = 7.15$. ■

Ejemplo 3.18 Sea $X = 1$ si un vehículo seleccionado al azar aprueba un diagnóstico de emisiones y $X = 0$ si no. Entonces X es una variable aleatoria de Bernoulli con función de masa de probabilidad $p(1) = p$ y $p(0) = 1 - p$ a partir de la cual $E(X) = 0 \cdot p(0) + 1 \cdot p(1) = 0(1 - p) + 1(p) = p$. Es decir, el valor esperado de X es exactamente la probabilidad de que X tome el valor de 1. Si se conceptualiza una población compuesta de ceros en la proporción $1 - p$ y números 1 en la proporción p, entonces el promedio de la población es $\mu = p$. ■

Ejemplo 3.19 La forma general de la función de masa de probabilidad de $X =$ número de niños nacidos hasta el primer varón incluido éste

$$p(x) = \begin{cases} p(1 - p)^{x-1} & x = 1, 2, 3, \ldots \\ 0 & \text{de lo contrario} \end{cases}$$

De acuerdo con la definición,

$$E(X) = \sum_D x \cdot p(x) = \sum_{x=1}^{\infty} xp(1 - p)^{x-1} = p\sum_{x=1}^{\infty}\left[- \frac{d}{dp}(1 - p)^x\right] \qquad (3.9)$$

Si se intercambia el orden en que se evalúan la derivada y la suma, la suma es la de una serie geométrica. Una vez que se calcula la suma, se saca la derivada y el resultado final es $E(X) = 1/p$. Si p se aproxima a 1, se espera ver que nazca un varón muy pronto, mientras que si p se aproxima a 0, se esperan muchos nacimientos antes del primer varón. Con $p = .5$, $E(X) = 2$. ■

Existe otra interpretación frecuentemente utilizada de μ. Considere la posibilidad de observar un primer valor x_1 de X, a continuación, un segundo valor x_2, un tercer valor x_3 y así sucesivamente. Después de hacer esto un gran número de veces, se calcula el promedio de la muestra de las x_i observadas. Este promedio será típicamente muy cercano a μ. Es decir, μ puede interpretarse como el promedio a largo plazo del valor observado de X cuando el experimento se realiza en varias ocasiones. En el ejemplo 3.17, el promedio a largo plazo de Apgar es $\mu = 7.15$.

Ejemplo 3.20 Sea X el número de entrevistas que un estudiante sostiene antes de conseguir un trabajo, y tiene la función de masa de probabilidad

$$p(x) = \begin{cases} k/x^2 & x = 1, 2, 3, \ldots \\ 0 & \text{de lo contrario} \end{cases}$$

donde k se elige de modo que $\sum_{x=1}^{\infty}(k/x^2) = 1$. (En un curso de matemáticas de series infinitas se demuestra que $\sum_{x=1}^{\infty}(1/x^2) < \infty$, lo cual implica que tal k existe, pero su valor exacto no interesa.) El valor esperado de X es

$$\mu = E(X) = \sum_{x=1}^{\infty} x \cdot \frac{k}{x^2} = k\sum_{x=1}^{\infty}\frac{1}{x} \qquad (3.10)$$

La suma del lado derecho de la ecuación (3.10) es la famosa serie armónica de matemáticas y se puede demostrar que es igual a ∞. $E(X)$ no es finita en este caso porque $p(x)$ no disminuye suficientemente rápido a medida que x se incrementa; los estadísticos dicen que la distribución de probabilidad de X tiene "una cola gruesa". Si se selecciona una secuencia de valores X utilizando esta distribución, el promedio muestral no se establecerá en un número finito sino que tenderá a crecer sin límite.

Los estadísticos utilizan la frase "colas gruesas" en conexión con cualquier distribución con una gran cantidad de probabilidad alejada de μ (así que las colas gruesas no requieren $\mu = \infty$). Tales colas gruesas hacen difícil hacer inferencias sobre μ. ■

Valor esperado de una función

A menudo interesará poner atención al valor esperado de alguna función $h(X)$ en lugar de sólo en $E(X)$.

Ejemplo 3.21 Suponga que una librería adquiere diez ejemplares de un libro a \$6.00 cada uno para venderlos a \$12.00 en el entendimiento de que al final de un periodo de 3 meses cualquier ejemplar no vendido puede ser compensado por \$2.00. Si X = el número de ejemplares vendidos, entonces el ingreso neto $= h(X) = 12X + 2(10 - X) - 60 = 10X - 40$. ¿Cuál es entonces el ingreso neto esperado? ■

El siguiente ejemplo sugiere una forma fácil de calcular el valor esperado de $h(X)$.

Ejemplo 3.22 El costo de cierta prueba de diagnóstico de un vehículo depende del número de cilindros X en el motor. Supóngase que la función de costo está dada por $h(X) = 20 + 3X + .5X^2$. Como X es una variable aleatoria, también lo es $Y = h(X)$. Las funciones de masa de probabilidad de X y Y son las siguientes

x	4	6	8	$\Rightarrow$	y	40	56	76
$p(x)$	.5	.3	.2		$p(y)$	.5	.3	.2

Con D^* denotando posibles valores de Y,

$$\begin{aligned} E(Y) = E[h(X)] &= \sum_{D^*} y \cdot p(y) \\ &= (40)(.5) + (56)(.3) + (76)(.2) \\ &= h(4) \cdot (.5) + h(6) \cdot (.3) + h(8) \cdot (.2) \\ &= \sum_{D} h(x) \cdot p(x) \end{aligned} \qquad (3.11)$$

De acuerdo con la ecuación (3.11), no fue necesario determinar la función de masa de probabilidad de Y para obtener $E(Y)$; en su lugar, el valor esperado deseado es un promedio ponderado de los posibles valores de $h(x)$ (y no de x). ■

PROPOSICIÓN

Si la variable aleatoria X tiene un conjunto de posibles valores D y una función de masa de probabilidad $p(x)$, entonces el valor esperado de cualquier función $h(X)$, denotada por $E[h(X)]$ o $\mu_{h(X)}$, se calcula con

$$E[h(X)] = \sum_{D} h(x) \cdot p(x)$$

Esto es, $E[h(X)]$ se calcula del mismo modo que $E(X)$, excepto que $h(x)$ sustituye a x.

Ejemplo 3.23 Una tienda de computadoras adquirió tres computadoras de un tipo a \$500 cada una. Las venderá a \$1000 cada una. El fabricante se comprometió a readquirir cualquier computadora que no se haya vendido después de un periodo especificado a \$200 cada una. Sea X el número de computadoras vendidas y suponga que $p(0) = .1$, $p(1) = .2$, $p(2) = .3$ y $p(3) = .4$. Con $h(X)$ denotando la utilidad asociada con la venta de X unidades, la información dada implica que $h(X) =$ ingreso $-$ costo $= 1000X + 200(3 - X) - 1500 = 800X - 900$. La utilidad esperada es entonces

$$\begin{aligned} E[h(X)] &= h(0) \cdot p(0) + h(1) \cdot p(1) + h(2) \cdot p(2) + h(3) \cdot p(3) \\ &= (-900)(.1) + (-100)(.2) + (700)(.3) + (1500)(.4) \\ &= \$700 \end{aligned}$$

■

Reglas de valor esperado

La función de interés $h(X)$ es con bastante frecuencia una función lineal $aX + b$. En este caso, $E[h(X)]$ es fácil de calcular a partir de $E(X)$.

PROPOSICIÓN

$$E(aX + b) = a \cdot E(X) + b$$

(O, con notación alternativa, $\mu_{aX+b} = a \cdot \mu_X + b$)

Parafraseando, el valor esperado de una función lineal es igual a la función lineal evaluada con el valor esperado $E(X)$. Como $h(X)$ en el ejemplo 3.23 es lineal y $E(X) = 2$, $E[h(x)] = 800(2) - 900 = \$700$, como antes.

Comprobación

$$\begin{aligned} E(aX + b) &= \sum_D (ax + b) \cdot p(x) = a\sum_D x \cdot p(x) + b\sum_D p(x) \\ &= aE(X) + b \end{aligned}$$

■

Dos casos especiales de proposición producen dos reglas importantes de valor esperado.

1. Con cualquier constante a, $E(aX) = a \cdot E(X)$ (considérese $b = 0$).
2. Con cualquier constante b, $E(X + b) = E(X) + b$ (considérese $a = 1$). (3.12)

La multiplicación de X por una constante a cambia por lo general la unidad de medición; por ejemplo, de pulgadas a centímetros, donde $a = 2.54$, etc.). La regla 1 dice que el valor esperado en las nuevas unidades es igual al valor esperado en las viejas unidades multiplicado por el factor de conversión a. Asimismo, si se agrega una constante b a cada valor posible de X, entonces el valor esperado se desplazará en esa misma cantidad constante.

Varianza de X

El valor esperado de X describe dónde está centrada la distribución de probabilidad. Utilizando la analogía física de colocar una masa puntual $p(x)$ en el valor x sobre un eje unidimensional que estuviera soportado por un fulcro colocado en μ, el eje no tendería a ladearse. Esto se ilustra para dos distribuciones diferentes en la figura 3.7.

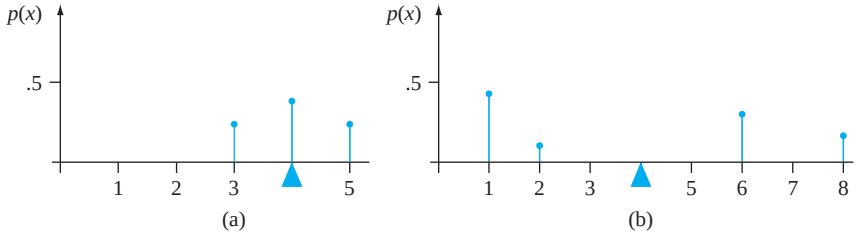


Figura 3.7 Dos diferentes distribuciones de probabilidad con $\mu = 4$

Aunque ambas distribuciones ilustradas en la figura 3.7 tienen el mismo centro μ, la distribución de la figura 3.7(b) tiene una mayor dispersión o variabilidad que la de la figura 3.7(a). Se utilizará la varianza de X para evaluar la cantidad de variabilidad en (la distribución de) X, del mismo modo que se utilizó s^2 en el capítulo 1 para medir la variabilidad en una muestra.

DEFINICIÓN

Sea $p(x)$ la función de masa de probabilidad de X y μ su valor esperado. En ese caso la **varianza** de X, denotada por $V(X)$ o σ_X^2, o simplemente σ^2, es

$$V(X) = \sum_D (x - \mu)^2 \cdot p(x) = E[(X - \mu)^2]$$

La **desviación estándar** (DE) de X es

$$\sigma_X = \sqrt{\sigma_X^2}$$

La cantidad $h(X) = (X - \mu)^2$ es la desviación al cuadrado de X con respecto a su media y σ^2 es la desviación al cuadrado esperada, es decir, el promedio ponderado de las desviaciones al cuadrado, donde las ponderaciones son probabilidades de la distribución. Si la mayor parte de la distribución de probabilidad está cerca de μ, entonces σ^2 será relativamente pequeña. Sin embargo, si existen valores x alejados de μ que tienen una gran $p(x)$, en ese caso σ^2 será bastante grande. A grandes rasgos, σ se puede interpretar como el tamaño de una desviación representativa del valor medio μ. Así que si $\sigma = 10$, entonces en una larga secuencia de valores observados X, algunos se apartarán de μ por más de 10, mientras que otros estarán más cerca de la media que eso, una desviación típica de la media será del orden de 10.

Ejemplo 3.24 Una biblioteca tiene un límite superior de 6 en el número de videos que puede sacar una persona a la vez. Tenga en cuenta sólo a quienes echan un vistazo a los videos y deje que X denote el número de videos que pide prestados para una persona seleccionada al azar. La función de masa de probabilidad de X es la siguiente:

x	1	2	3	4	5	6
$p(x)$	.30	.25	.15	.05	.10	.15

Es fácil ver que el valor esperado de X es $\mu = 2.85$. La varianza de X es entonces

$$V(X) = \sigma^2 = \sum_{x=1}^{6} (x - 2.85)^2 \cdot p(x)$$
$$= (1 - 2.85)^2(.30) + (2 - 2.85)^2(.25) + \cdots + (6 - 2.85)^2(.15) = 3.2275$$

La desviación estándar de X es $\sigma = \sqrt{3.2275} = 1.800$. ■

Cuando la función de masa de probabilidad $p(x)$ especifica un modelo matemático para la distribución de los valores de la población, tanto σ^2 como σ miden la dispersión de los valores en la población; σ^2 es la varianza de la población y σ es su desviación estándar.

Fórmula abreviada para σ^2

El número de operaciones aritméticas necesarias para calcular σ^2 puede reducirse si se utiliza una fórmula alternativa.

PROPOSICIÓN

$$V(X) = \sigma^2 = \left[\sum_D x^2 \cdot p(x)\right] - \mu^2 = E(X^2) - [E(X)]^2$$

Al utilizar esta fórmula, $E(X^2)$ se calcula primero sin ninguna sustracción; acto seguido $E(X)$ se calcula, se eleva al cuadrado y se resta (una vez) de $E(X^2)$.

Ejemplo 3.25 (Continuación del ejemplo 3.24)

La función de masa de probabilidad del número X de videos prestados se dio como $p(1) = .30$, $p(2) = .25$, $p(3) = .15$, $p(4) = .05$, $p(5) = .10$ y $p(6) = .15$, a partir de las cuales $\mu = 2.85$ y

$$E(X^2) = \sum_{x=1}^{6} x^2 \cdot p(x) = (1^2)(.30) + (2^2)(.25) + \cdots + (6^2)(.15) = 11.35$$

Por lo tanto, $\sigma^2 = 11.35 - (2.85)^2 = 3.2275$, como se obtuvo previamente de la definición. ■

Demostración de la fórmula abreviada Desarróllese $(x - \mu)^2$ en la definición de σ^2 para obtener $x^2 - 2\mu x + \mu^2$, y luego lleve $\sum$ a cada uno de los tres términos:

$$\begin{aligned}\sigma^2 &= \sum_D x^2 \cdot p(x) - 2\mu \cdot \sum_D x \cdot p(x) + \mu^2 \sum_D p(x) \\ &= E(X^2) - 2\mu \cdot \mu + \mu^2 = E(X^2) - \mu^2\end{aligned}$$

■

Reglas de varianza

La varianza de $h(X)$ es el valor esperado de la diferencia al cuadrado entre $h(X)$ y su valor esperado:

$$V[h(X)] = \sigma^2_{h(X)} = \sum_D \{h(x) - E[h(X)]\}^2 \cdot p(x) \tag{3.13}$$

Cuando $h(X) = aX + b$, una función lineal,

$$h(x) - E[h(X)] = ax + b - (a\mu + b) = a(x - \mu)$$

Sustituyendo esto en la ecuación (3.13) se obtiene una relación simple entre $V[h(X)]$ y $V(X)$:

PROPOSICIÓN

$$V(aX + b) = \sigma^2_{aX+b} = a^2 \cdot \sigma^2_X \text{ y } \sigma_{aX+b} = |a| \cdot \sigma_x$$

En particular,

$$\sigma_{aX} = |a| \cdot \sigma_X, \quad \sigma_{X+b} = \sigma_X \tag{3.14}$$

El valor absoluto es necesario porque a podría ser negativa, no obstante una desviación estándar no puede serlo. Casi siempre la multiplicación por a corresponde a un cambio de la unidad de medición (p. ej., kg a lb o dólares a euros). De acuerdo con la primera relación en (3.14), la desviación estándar en la nueva unidad es la desviación estándar original multiplicada por el factor de conversión. La segunda relación dice que la adición o sustracción de una constante no impacta la variabilidad; simplemente desplaza la distribución a la derecha o izquierda.

Ejemplo 3.26 En el problema de ventas de computadoras del ejemplo 3.23, $E(X) = 2$ y

$$E(X^2) = (0)^2(.1) + (1)^2(.2) + (2)^2(.3) + (3)^2(.4) = 5$$

así que $V(X) = 5 - (2)^2 = 1$. La función de utilidad $h(X) = 800X - 900$ tiene entonces la varianza $(800)^2 \cdot V(X) = (640{,}000)(1) = 640{,}000$ y la desviación estándar 800. ■

EJERCICIOS Sección 3.3 (29–45)

29. La función de masa de probabilidad de la cantidad de memoria X (GB) en una unidad flash comprada fue dada en el ejemplo 3.13 como

x	1	2	4	8	16
$p(x)$	.05	.10	.35	.40	.10

Calcule lo siguiente:
a. $E(X)$
b. $V(X)$ directamente a partir de la definición
c. La desviación estándar de X
d. $V(X)$ por medio de la fórmula abreviada

30. Se selecciona al azar un individuo que tiene asegurado su automóvil con una compañía. Sea Y el número de infracciones de tránsito por las que el individuo fue citado durante los últimos tres años. La función de masa de probabilidad de Y es

y	0	1	2	3
$p(y)$	.60	.25	.10	.05

a. Calcule $E(Y)$.
b. Suponga que un individuo con Y infracciones incurre en un recargo de $\$100Y^2$. Calcule el monto esperado del recargo.

31. Remítase al ejercicio 12 y calcule $V(Y)$ y σ_Y. Determine entonces la probabilidad de que Y esté dentro de una desviación estándar de 1 de su valor medio.

32. Un distribuidor de enseres para el hogar vende tres modelos de congeladores verticales de 13.5, 15.9 y 19.1 pies cúbicos de espacio de almacenamiento, respectivamente. Sea X = la cantidad de espacio de almacenamiento adquirido por el siguiente cliente que compre un congelador. Suponga que X tiene la función de masa de probabilidad

x	13.5	15.9	19.1
$p(x)$	.2	.5	.3

a. Calcule $E(X)$, $E(X^2)$ y $V(X)$.
b. Si el precio de un congelador de X pies cúbicos de capacidad es $25X - 8.5$, ¿cuál es el precio esperado pagado por el siguiente cliente que compre un congelador?
c. ¿Cuál es la varianza del precio $25X - 8.5$ pagado por el siguiente cliente?
d. Suponga que aunque la capacidad nominal de un congelador es X, la real es $h(X) = X - .01X^2$. ¿Cuál es la capacidad real esperada del congelador adquirido por el siguiente cliente?

33. Sea X una variable aleatoria de Bernoulli con función de masa de probabilidad como en el ejemplo 3.18.
a. Calcule $E(X^2)$
b. Demuestre que $V(X) = p(1 - p)$.
c. Calcule $E(X^{79})$.

34. Suponga que el número de plantas de un tipo particular encontradas en una región rectangular de muestreo (llamada cuadrado por los ecologistas) en cierta área geográfica es una variable aleatoria X con función de masa de probabilidad

$$p(x) = \begin{cases} c/x^3 & x = 1, 2, 3, \ldots \\ 0 & \text{de lo contrario} \end{cases}$$

¿Es $E(X)$ finita? Justifique su respuesta (ésta es otra distribución que los estadísticos llamarían de cola gruesa).

35. Un pequeño mercado ordena ejemplares de cierta revista para su exhibidor de revistas cada semana. Sea X = demanda de la revista, con función de masa de probabilidad

x	1	2	3	4	5	6
$p(x)$	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{3}{15}$	$\frac{4}{15}$	$\frac{3}{15}$	$\frac{2}{15}$

Suponga que el propietario de la tienda paga $2.00 por cada ejemplar de la revista y el precio para los consumidores es de $4.00. Si las revistas que se quedan al final de la semana no tienen valor de recuperación, ¿es mejor ordenar tres o cuatro ejemplares de la revista? [*Sugerencia*: para tres o cuatro ejemplares ordenados, exprese el ingreso neto como una función de la demanda X, y luego calcule el ingreso esperado.]

36. Sea X el daño incurrido (en dólares) en un tipo de accidente durante un año dado. Valores posibles de X son 0, 1000, 5000 y 10,000, con probabilidades de .8, .1, .08 y .02, respectivamente. Una compañía particular ofrece una póliza con deducible de $500. Si la compañía desea que su utilidad esperada sea de $100, ¿qué cantidad de prima deberá cobrar?

37. Los n candidatos para un trabajo fueron clasificados como 1, 2, 3, . . . , n. Sea X = la clasificación de un candidato seleccionado al azar, de modo que X tenga la función de masa de probabilidad

$$p(x) = \begin{cases} 1/n & x = 1, 2, 3, \ldots, n \\ 0 & \text{de lo contrario} \end{cases}$$

(ésta se llama *distribución uniforme discreta*). Calcule $E(X)$ y $V(X)$ por medio de la fórmula abreviada. [*Sugerencia*: la suma de los primeros n enteros positivos es $n(n + 1)/2$, mientras que la suma de sus cuadrados es $n(n + 1)(2n + 1)/6$.]

38. Sea X = el resultado cuando un dado imparcial es lanzado una vez. Si antes de lanzar el dado le ofrecen (1/3.5) dólares o $h(X) = 1/X$ dólares, ¿aceptaría la suma garantizada o jugaría? [*Nota*: generalmente no es cierto que $1/E(X) = E(1/X)$.]

39. Una compañía de productos químicos en la actualidad tiene en existencia 100 lb de un producto químico, el cual se vende a sus clientes en lotes de 5 lb. Sea X = el número de lotes solicitados por un cliente seleccionado al azar y suponga que X tiene la función de masa de probabilidad

x	1	2	3	4
$p(x)$	.2	.4	.3	.1

Calcule $E(X)$ y $V(X)$. Calcule en seguida el número esperado de libras que quedan una vez que se envía el pedido del siguiente cliente y la varianza del número de libras que quedan. [*Sugerencia*: el número de libras que quedan es una función lineal de X.]

40. **a.** Trace una gráfica lineal de la función de masa de probabilidad de X en el ejercicio 35. En seguida determine la función de masa de probabilidad de $-X$ y trace su gráfica lineal. Con base en estas dos figuras, ¿qué se puede decir sobre $V(X)$ y $V(-X)$?

b. Use la proposición que implica $V(aX + b)$ para establecer una relación general entre $V(X)$ y $V(-X)$.

41. Use la definición en la expresión (3.13) para comprobar que $V(aX + b) = a^2 \cdot \sigma_X^2$. [*Sugerencia*: con $h(X) = aX + b$, $E[h(X)] = a\mu + b$, donde $\mu = E(X)$.]

42. Suponga $E(X) = 5$ y $E[X(X - 1)] = 27.5$. ¿Cuál es

a. $E(X^2)$? [*Sugerencia*: $E[X(X - 1)] = E(X^2 - X] = E(X^2) - E(X)$]?

b. $V(X)$?

c. la relación general entre las cantidades $E(X)$, $E[X(X) - 1)]$ y $V(X)$?

43. Escriba una regla general para $E(X - c)$, donde c es una constante. ¿Qué sucede cuando hace $c = \mu$, el valor esperado de X?

44. Un resultado llamado **desigualdad de Chebyshev** establece que para cualquier distribución de probabilidad de una variable aleatoria X y cualquier número k que por lo menos sea 1, $P(|X - \mu| \geq k\sigma) \leq 1/k^2$. En palabras, la posibilidad de que el valor de X quede por lo menos a k desviaciones estándar de su media es cuando mucho $1/k^2$.

a. ¿Cuál es el valor del límite superior con $k = 2$? ¿$k = 3$? ¿$k = 4$? ¿$k = 5$? ¿$k = 10$?

b. Calcule μ y σ para la distribución del ejercicio 13. Evalúe en seguida $P(|X - \mu| \geq k\sigma)$ con los valores de k dados en el inciso (a). ¿Qué sugiere esto sobre el límite superior con respecto a la probabilidad correspondiente?

c. Sea X con los valores posibles -1, 0 y 1, con las probabilidades $\frac{1}{18}$, $\frac{8}{9}$ y $\frac{1}{18}$, respectivamente. ¿Cuál es $P(|X - \mu| \geq 3\sigma)$ y cómo se compara con el límite correspondiente?

d. Dé una distribución para la cual $P(|X - \mu| \geq 5\sigma) = .04$.

45. Si $a \leq X \leq b$, demuestre que $a \leq E(X) \leq b$.

3.4 Distribución de probabilidad binomial

Existen muchos experimentos que se ajustan exacta o aproximadamente a la siguiente lista de requerimientos.

1. El experimento consta de una secuencia de n experimentos más pequeños llamados *ensayos*, donde n se fija antes del experimento.
2. Cada ensayo puede dar por resultado uno de los mismos dos resultados posibles (ensayos dicotómicos), los cuales se denotan como éxito (S) y falla (F).
3. Los ensayos son independientes, de modo que el resultado en cualquier ensayo particular no influye en el resultado de cualquier otro ensayo.
4. La probabilidad de éxito $P(S)$ es constante de un ensayo a otro; esta probabilidad se denota por p.

DEFINICIÓN

Un experimento para el que se satisfacen las condiciones 1–4 se llama **experimento binomial.**

Ejemplo 3.27 La misma moneda se lanza al aire sucesiva e independientemente n veces. De manera arbitraria se utiliza S para denotar el resultado H (caras) y F para denotar el resultado T (cruces). Entonces este experimento satisface las condiciones 1–4. El lanzamiento al aire de una tachuela n veces, con S = punta hacia arriba y F = punta hacia abajo, también da por resultado un experimento binomial. ■

Muchos experimentos implican una secuencia de ensayos independientes para los cuales existen más de dos resultados posibles en cualquier ensayo. Entonces, un experimento binomial puede crearse dividiendo los posibles resultados en dos grupos.

Ejemplo 3.28 El color de las semillas de chícharo lo determina un solo locus genético. Si los dos alelos en este locus genético son AA o Aa (el genotipo), entonces el chícharo será amarillo (el fenotipo) y si el alelo es aa, el chícharo será verde. Suponga que se aparean 20 semillas Aa y se cruzan las dos semillas en cada uno de los diez pares para obtener diez nuevos genotipos. Designe a cuada nuevo genotipo como éxito (S) si es aa y falla (F) si es lo contrario. Entonces con esta identificación de S y F, el experimento es binomial con $n = 10$ y $p = P$ (genotipo aa). Si es igualmente probable que cada miembro del par contribuya con a o A, entonces $p = P(a) \cdot P(a) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4}$. ■

Ejemplo 3.29 Suponga que una ciudad tiene 50 restaurantes autorizados, de los cuales 15 han cometido en la actualidad una seria violación del código sanitario y los otros 35 no han cometido violaciones serias. Hay cinco inspectores, cada uno de los cuales inspeccionará un restaurante durante la semana entrante. El nombre de cada restaurante se anota en un pedacito de papel diferente y a continuación se mezclan perfectamente, cada inspector a su vez saca uno de los papelitos *sin reemplazarlos*. Anótese el ensayo i-ésimo como éxito si el restaurante i-ésimo seleccionado ($i = 1, \ldots, 5$) no ha cometido violaciones serias. Entonces

$$P(S \text{ en el primer ensayo}) = \frac{35}{50} = .70$$

y

$$\begin{aligned} P(S \text{ en el segundo ensayo}) &= P(SS) + P(FS) \\ &= P(\text{segundo } S \mid \text{primer } S)P(\text{primer } S) \\ &\quad + P(\text{segundo } S \mid \text{primera } F)P(\text{primera } F) \\ &= \frac{34}{49} \cdot \frac{35}{50} + \frac{35}{49} \cdot \frac{15}{50} = \frac{35}{50}\left(\frac{34}{49} + \frac{15}{49}\right) = \frac{35}{50} = .70 \end{aligned}$$

De manera similar, se puede demostrar que $P(S \text{ en el ensayo } i\text{-ésimo}) = .70$ con $i = 3, 4, 5$. Sin embargo,

$$P(S \text{ en el quinto ensayo} \mid SSSS) = \frac{31}{46} = .67$$

mientras que

$$P(S \text{ en el quinto ensayo} \mid FFFF) = \frac{35}{46} = .76$$

El experimento no es binomial porque los ensayos no son independientes. En general, si se muestrea sin reemplazo, el experimento no producirá ensayos independientes. Si cada papelito hubiera sido reemplazado después de ser sacado, entonces los ensayos habrían sido independientes, pero esto podría haber dado por resultado que el mismo restaurante fuera inspeccionado por más de un inspector. ■

Ejemplo 3.30 Un estado tiene 500,000 conductores con licencia, de los cuales 400,000 están asegurados. Se selecciona una muestra de 10 conductores sin reemplazo. El ensayo *i*-ésimo se denota *S* si el conductor *i*ésimo seleccionado está asegurado. Aunque esta situación parecería idéntica a la del ejemplo 3.29, la diferencia importante es que el tamaño de la población muestreada es muy grande con respecto al tamaño de la muestra. En este caso

$$P(S \text{ en } 2 \mid S \text{ en } 1) = \frac{399{,}999}{499{,}999} = .80000$$

y

$$P(S \text{ en } 10 \mid S \text{ en los primeros } 9) = \frac{399{,}991}{499{,}991} = .799996 \approx .80000$$

Estos cálculos sugieren que aunque los ensayos no son exactamente independientes, las probabilidades condicionales difieren tan poco una de otra que para propósitos prácticos los ensayos se consideran independientes con la constante $P(S) = .8$. Por lo tanto, para una muy buena aproximación, el experimento es binomial con $n = 10$ y $p = .8$. ■

Se utilizará la siguiente regla empírica para decidir si un experimento "sin reemplazo" puede ser tratado como un experimento binomial.

REGLA

Considérese muestreo sin reemplazo de una población dicotómica de tamaño N. Si el tamaño de la muestra (número de ensayos) n es cuando mucho 5% del tamaño de la población, el experimento puede ser analizado como si fuera exactamente un experimento binomial.

Por "analizado" se quiere decir que las probabilidades basadas en suposiciones de experimento binomial se aproximarán bastante a las probabilidades reales "sin reemplazo", las que generalmente son más difíciles de calcular. En el ejemplo 3.29, $n/N = 5/50 = .1 > .05$, de modo que el experimento binomial no es una buena aproximación, pero en el ejemplo 3.30, $n/N = 10/500{,}000 < .05$.

Variable y distribución aleatoria binomial

En la mayoría de los experimentos binomiales, lo que interesa es el número total de los éxitos (S), en lugar del conocimiento de qué ensayos dieron los éxitos.

DEFINICIÓN

La **variable aleatoria binomial *X*** asociada con un experimento binomial que consiste en n ensayos se define como

X = el número de los S entre los n ensayos

Supóngase, por ejemplo, que $n = 3$. Entonces existen ocho posibles resultados para el experimento:

SSS *SSF* *SFS* *SFF* *FSS* *FSF* *FFS* *FFF*

Por la definición de X, $X(SSF) = 2$, $X(SFF) = 1$, y así sucesivamente. Valores posibles de X en un experimento de n ensayos son $x = 0, 1, 2, \ldots, n$. A menudo se escribirá $X \sim \text{Bin}(n, p)$ para indicar que X es una variable aleatoria binomial basada en n ensayos con probabilidad de éxito p.

NOTACIÓN

Como la función de masa de probabilidad de una variable aleatoria binomial X depende de los dos parámetros n y p, la función de masa de probabilidad se denota por $b(x; n, p)$.

Considérese primero el caso $n = 4$ para el cual cada resultado, su probabilidad y su valor x correspondiente se dan en la tabla 3.1. Por ejemplo,

$$\begin{aligned} P(SSFS) &= P(S) \cdot P(S) \cdot P(F) \cdot P(S) \text{ (ensayos independientes)} \\ &= p \cdot p \cdot (1 - p) \cdot p \text{ (constante } P(S)) \\ &= p^3 \cdot (1 - p) \end{aligned}$$

Tabla 3.1 Resultados y probabilidades para un experimento binomial con cuatro intentos

Resultado	x	Probabilidad	Resultado	x	Probabilidad
SSSS	4	p^4	*FSSS*	3	$p^3(1 - p)$
SSSF	3	$p^3(1 - p)$	*FSSF*	2	$p^2(1 - p)^2$
SSFS	3	$p^3(1 - p)$	*FSFS*	2	$p^2(1 - p)^2$
SSFF	2	$p^2(1 - p)^2$	*FSFF*	1	$p(1 - p)^3$
SFSS	3	$p^3(1 - p)$	*FFSS*	2	$p^2(1 - p)^2$
SFSF	2	$p^2(1 - p)^2$	*FFSF*	1	$p(1 - p)^3$
SFFS	2	$p^2(1 - p)^2$	*FFFS*	1	$p(1 - p)^3$
SFFF	1	$p(1 - p)^3$	*FFFF*	0	$(1 - p)^4$

En este caso especial, se desea $b(x; 4, p)$ con $x = 0, 1, 2, 3$ y 4. Para $b(3; 4, p)$, identifíquese cuáles de los 16 resultados dan un valor x de 3 y sume las probabilidades asociadas con cada resultado:

$$b(3; 4, p) = P(FSSS) + P(SFSS) + P(SSFS) + P(SSSF) = 4p^3(1 - p)$$

Existen cuatro resultados con $X = 3$ y la probabilidad de cada uno es $p^3(1 - p)$(el orden de los S y las F no es importante, sino sólo el número de los S), por lo tanto

$$b(3; 4, p) = \left\{\begin{matrix}\text{número de resultados} \\ \text{con } X = 3\end{matrix}\right\} \cdot \left\{\begin{matrix}\text{probabilidad de cualquier} \\ \text{resultado con } X = 3\end{matrix}\right\}$$

Asimismo, $b(2; 4, p) = 6p^2(1 - p)^2$, la cual también es el producto del número de resultados con $X = 2$ y la probabilidad de cualquier resultado como ése.

En general,

$$b(x; n, p) = \left\{\begin{matrix}\text{número de secuencias de longitud } n \\ \text{compuestas de } x \text{ éxitos}\end{matrix}\right\} \cdot \left\{\begin{matrix}\text{probabilidad de cualquier} \\ \text{secuencia como ésa}\end{matrix}\right\}$$

Como el orden de los S y las F no es importante, el segundo factor en la ecuación previa es $p^x(1 - p)^{n-x}$ (p. ej., los primeros x ensayos producen S y los últimos $n - x$ producen F). El primer factor es el número de formas de escoger x de los n ensayos para que sean los S, es decir, el número de combinaciones de tamaño x que pueden ser construidas con n objetos distintos (ensayos en este caso).

TEOREMA

$$b(x; n, p) = \begin{cases} \binom{n}{x} p^x(1 - p)^{n-x} & x = 0, 1, 2, \ldots, n \\ 0 & \text{de lo contrario} \end{cases}$$

Ejemplo 3.31 A cada uno de seis bebedores de refrescos de cola seleccionados al azar se le sirve un vaso de refresco de cola *S* y uno de refresco de cola *F*. Los vasos son idénticos en apariencia excepto por un código que viene en el fondo para identificar el refresco de cola. Suponga que en realidad no existe una tendencia entre los bebedores de refresco de cola de preferir un refresco de cola en vez del otro. Entonces $p = P(\text{un individuo seleccionado prefiere } S) = .5$, así que con $X =$ el número entre los seis que prefieren *S*, $X \sim \text{Bin}(6, .5)$.

Por lo tanto

$$P(X = 3) = b(3; 6, .5) = \binom{6}{3}(.5)^3(.5)^3 = 20(.5)^6 = .313$$

La probabilidad de que por lo menos tres prefieran *S* es

$$P(3 \le X) = \sum_{x=3}^{6} b(x; 6, .5) = \sum_{x=3}^{6} \binom{6}{x}(.5)^x(.5)^{6-x} = .656$$

y la probabilidad de que cuando mucho uno prefiera *S* es

$$P(X \le 1) = \sum_{x=0}^{1} b(x; 6, .5) = .109$$ ■

Utilización de tablas binomiales*

Incluso con un valor relativamente pequeño de *n*, el cálculo de probabilidades binomiales es tedioso. La tabla A.1 del apéndice tabula la función de distribución acumulativa $F(x) = P(X \le x)$ con $n = 5, 10, 15, 20, 25$ en combinación con valores seleccionados de *p*. Varias otras probabilidades pueden entonces ser calculadas por medio de la proposición sobre funciones de distribución acumulativa de la sección 3.2. Una anotación de 0 en la tabla significa únicamente que la probabilidad es 0 a tres dígitos significativos puesto que todos los valores ingresados en la tabla en realidad son positivos.

NOTACIÓN

Para $X \sim \text{Bin}(n, p)$, la función de distribución acumulativa será denotada por

$$B(x; n, p) = P(X \le x) = \sum_{y=0}^{x} b(y; n, p) \quad x = 0, 1, \ldots, n$$

Ejemplo 3.32 Suponga que 20% de todos los ejemplares de un libro de texto particular no pasan una prueba de resistencia de encuadernación. Sea *X* el número entre 15 ejemplares seleccionados al azar que no pasan la prueba. Entonces *X* tiene una distribución binomial con $n = 15$ y $p = .2$.

1. La probabilidad de que cuando mucho 8 no pasen la prueba es

$$P(X \le 8) = \sum_{y=0}^{8} b(y; 15, .2) = B(8; 15, .2)$$

la cual es el dato en el renglón $x = 8$ y la columna $p = .2$ de la tabla binomial $n = 15$. Según la tabla A.1 del apéndice, la probabilidad es $B(8; 15, .2) = .999$.

2. La probabilidad de que exactamente 8 fallen es

$$P(X = 8) = P(X \le 8) - P(X \le 7) = B(8; 15, .2) - B(7; 15, .2)$$

la cual es la diferencia entre dos datos consecutivos en la columna $p = .2$. El resultado es $.999 - .996 = .003$.

* Los paquetes de programas estadísticos tales como Minitab y R proporcionan la función de masa de probabilidad o la función de distribución acumulativa en forma casi instantánea al solicitarla para cualquier valor de *p* y *n* desde 2 hasta millones. También existe un comando R para calcular la probabilidad de que *X* quede en algún intervalo.

3. La probabilidad de que por lo menos 8 fallen es

$$\begin{aligned} P(X \geq 8) &= 1 - P(X \leq 7) = 1 - B(7; 15, .2) \\ &= 1 - \begin{pmatrix} \text{dato en } x = 7 \\ \text{renglón de la columna } p = .2 \end{pmatrix} \\ &= 1 - .996 = .004 \end{aligned}$$

4. Finalmente, la probabilidad de que entre 4 y 7, inclusive, fallen es

$$\begin{aligned} P(4 \leq X \leq 7) &= P(X = 4, 5, 6 \text{ o } 7) = P(X \leq 7) - P(X \leq 3) \\ &= B(7; 15, .2) - B(3; 15, .2) = .996 - .648 = .348 \end{aligned}$$

Obsérvese que esta última probabilidad es la diferencia entre los datos en los renglones $x = 7$ y $x = 3$, *no* en los renglones $x = 7$ y $x = 4$. ■

Ejemplo 3.33 Un fabricante de aparatos electrónicos afirma que cuando mucho 10% de sus unidades de suministro de potencia necesitan servicio durante el periodo de garantía. Para investigar esta afirmación, técnicos en un laboratorio de prueba adquieren 20 unidades y someten a cada una a una prueba acelerada para simular el uso durante el periodo de garantía. Sea p la probabilidad de que una unidad de suministro de potencia necesite reparación durante el periodo (proporción de unidades que requieren reparación). Los técnicos de laboratorio deben decidir si los datos obtenidos con el experimento respaldan la afirmación de que $p \leq .10$. Sea X el número entre las 20 muestreadas que necesitan reparación, así que $X \sim \text{Bin}(20, p)$. Considere la regla de decisión:

Rechazar la afirmación de que $p \leq .10$ a favor de la conclusión de que $p > .10$ si $x \geq 5$ (donde x es el valor observado de X) y considere plausible la afirmación si $x \leq 4$.

La probabilidad de que la afirmación sea rechazada cuando $p = .10$ (una conclusión incorrecta) es

$$P(X \geq 5 \text{ cuando } p = .10) = 1 - B(4; 20, .1) = 1 - .957 = .043$$

La probabilidad de que la afirmación no sea rechazada cuando $p = .20$ (un tipo diferente de conclusión incorrecta) es

$$P(X \leq 4 \text{ cuando } p = .2) = B(4; 20, .2) = .630$$

La primera probabilidad es algo pequeña, pero la segunda es intolerablemente grande. Cuando $p = .20$, significa que el fabricante subestimó de manera excesiva el porcentaje de unidades que necesitan servicio, y si se utiliza la regla de decisión establecida, ¡el 63% de todas las muestras resultaron plausibles!

Se podría pensar que la probabilidad de este segundo tipo de conclusión errónea podría hacerse más pequeña cambiando el valor de corte de 5 en la regla de decisión a algo más. Sin embargo, aunque el reemplazo de 5 por un número más pequeño daría una probabilidad más pequeña que .630, la otra probabilidad se incrementaría entonces. La única forma de hacer ambas "probabilidades de error" pequeñas es basar la regla de decisión en un experimento que implique muchas más unidades. ■

La media y varianza de X

Con $n = 1$, la distribución binomial llega a ser la distribución de Bernoulli. De acuerdo con el ejemplo 3.18, el valor medio de una variable de Bernoulli es $\mu = p$, así que el número esperado de las S en cualquier ensayo único es p. Como un experimento binomial se compone de n ensayos, la intuición sugiere que para $X \sim \text{Bin}(n, p)$, $E(X) = np$, el producto del número de ensayos y la probabilidad de éxito en un solo ensayo. La expresión para $V(X)$ no es tan intuitiva.

PROPOSICIÓN

Si $X \sim \text{Bin}(n, p)$, entonces $E(X) = np$, $V(X) = np(1 - p) = npq$ y $\sigma_X = \sqrt{npq}$ (donde $q = 1 - p$).

Por tanto, para calcular la media y varianza de una variable aleatoria binomial no se requiere evaluar las sumas. La comprobación del resultado para $E(X)$ se ilustra en el ejercicio 64.

Ejemplo 3.34 Si 75% de todas las compras en una tienda se hacen con tarjeta de crédito y X es el número entre diez compras seleccionadas al azar realizadas con tarjeta de crédito, entonces $X \sim \text{Bin}(10, .75)$. Por lo tanto, $E(X) = np = (10)(.75) = 7.5$, $V(X) = npq = 10(.75)(.25) = 1.875$ y $\sigma = \sqrt{1.875} = 1.37$. Otra vez, aun cuando X puede tomar sólo valores enteros, $E(X)$ no tiene que ser un entero. Si se realiza un gran número de experimentos binomiales independientes, cada uno con $n = 10$ ensayos y $p = .75$, entonces el número promedio de las S por experimento se acercará a 7.5.

La probabilidad de que X se encuentre dentro de una desviación estándar de su valor medio es $P(7.5 - 1.37 \le X \le 7.5 + 1.37) = P(6.13 \le X \le 8.87) = P(X = 7 \text{ u } 8) = .532$. ■

EJERCICIOS Sección 3.4 (46–67)

46. Calcule las siguientes probabilidades binomiales directamente con la fórmula para $b(x; n, p)$:
- **a.** $b(3; 8, .35)$
- **b.** $b(5; 8, .6)$
- **c.** $P(3 \le X \le 5)$ cuando $n = 7$ y $p = .6$
- **d.** $P(1 \le X)$ cuando $n = 9$ y $p = .1$

47. Use la tabla A.1 del apéndice para obtener las siguientes probabilidades:
- **a.** $B(4; 15, .3)$
- **b.** $b(4; 15, .3)$
- **c.** $b(6; 15, .7)$
- **d.** $P(2 \le X \le 4)$ cuando $X \sim \text{Bin}(15, .3)$
- **e.** $P(2 \le X)$ cuando $X \sim \text{Bin}(15, .3)$
- **f.** $P(X \le 1)$ cuando $X \sim \text{Bin}(15, .7)$
- **g.** $P(2 < X < 6)$ cuando $X \sim \text{Bin}(15, .3)$

48. Cuando las tarjetas de circuito usadas en la fabricación de reproductores de discos compactos se prueban, el porcentaje de defectuosas es de 5%. Sea X = el número de tarjetas defectuosas en una muestra aleatoria de tamaño $n = 25$, así que $X \sim \text{Bin}(25, .05)$.
- **a.** Determine $P(X \le 2)$.
- **b.** Determine $P(X \ge 5)$.
- **c.** Determine $P(1 \le X \le 4)$.
- **d.** ¿Cuál es la probabilidad de que ninguna de estas 25 tarjetas esté defectuosa?
- **e.** Calcule el valor esperado y la desviación estándar de X.

49. Una compañía que produce cristal fino sabe por experiencia que 10% de sus copas de mesa tienen imperfecciones cosméticas y deben ser clasificadas como “segundas”.
- **a.** Entre seis copas seleccionadas al azar, ¿qué tan probable es que sólo una sea segunda?
- **b.** Entre seis copas seleccionadas al azar, ¿qué tan probable es que por lo menos dos sean segundas?
- **c.** Si las copas se examinan una por una, ¿cuál es la probabilidad de que cuando mucho cinco deban ser seleccionadas para encontrar cuatro que no sean segundas?

50. Se utiliza un número telefónico particular para recibir tanto llamadas de voz como faxes. Suponga que 25% de las llamadas entrantes son faxes y considere una muestra de 25 llamadas entrantes. ¿Cuál es la probabilidad de que
- **a.** cuando mucho 6 de las llamadas sean un fax?
- **b.** exactamente 6 de las llamadas sean un fax?
- **c.** por lo menos 6 de las llamadas sean un fax?
- **d.** más de 6 de las llamadas sean un fax?

51. Remítase al ejercicio previo.
- **a.** ¿Cuál es el número esperado de llamadas entre las 25 que implican un fax?
- **b.** ¿Cuál es la desviación estándar del número entre las 25 llamadas que implican un fax?
- **c.** ¿Cuál es la probabilidad de que el número de llamadas entre las 25 que implican una transmisión de fax sobrepase el número esperado por más de 2 desviaciones estándar?

52. Suponga que 30% de todos los estudiantes que tienen que comprar un texto para un curso particular desean un ejemplar nuevo (¡los exitosos!), mientras que el otro 70% desea comprar un ejemplar usado. Considere seleccionar 25 compradores al azar.
- **a.** ¿Cuáles son el valor medio y la desviación estándar del número que desea un ejemplar nuevo del libro?
- **b.** ¿Cuál es la probabilidad de que el número que desea ejemplares nuevos esté a más de dos desviaciones estándar del valor medio?

c. La librería tiene 15 ejemplares nuevos y 15 usados en existencia. Si 25 personas llegan una por una a comprar el texto, ¿cuál es la probabilidad de que las 25 obtengan el tipo de libro que desean de las existencias actuales? [*Sugerencia*: sea X = el número que desea un ejemplar nuevo. ¿Con qué valores de X obtendrán las 25 personas lo que desean?]

d. Suponga que los ejemplares nuevos cuestan \$100 y los usados \$70. Suponga que la librería en la actualidad tiene 50 ejemplares nuevos y 50 usados. ¿Cuál es el valor esperado del ingreso total por la venta de los siguientes 25 ejemplares comprados? Asegúrese de indicar qué regla de valor esperado está utilizando. [*Sugerencia*: sea $h(X)$ = el ingreso cuando X de los 25 compradores desean ejemplares nuevos. Exprese esto como una función lineal.]

53. El ejercicio 30 (sección 3.3) dio la función de masa de probabilidad de Y, el número de infracciones de tránsito de un individuo seleccionado al azar asegurado por una compañía particular. ¿Cuál es la probabilidad de que entre 15 individuos seleccionados al azar

a. por lo menos 10 no tengan infracciones?

b. menos de la mitad tengan por lo menos una infracción?

c. el número que tengan por lo menos una infracción esté entre 5 y 10, inclusive?*

54. Un tipo particular de raqueta de tenis viene en tamaño mediano y en tamaño extragrande. Sesenta por ciento de todos los clientes en una tienda desean la versión extragrande.

a. Entre diez clientes seleccionados al azar que desean este tipo de raqueta, ¿cuál es la probabilidad de que por lo menos seis deseen la versión extragrande?

b. Entre diez clientes seleccionados al azar, ¿cuál es la probabilidad de que el número que desea la versión extragrande esté dentro de 1 desviación estándar del valor medio?

c. La tienda dispone actualmente de siete raquetas de cada versión. ¿Cuál es la probabilidad de que los siguientes diez clientes que desean esta raqueta puedan obtener la versión que desean de las existencias actuales?

55. Veinte por ciento de todos los teléfonos de cierto tipo son llevados a servicio mientras se encuentran dentro de la garantía. De éstos, 60% pueden ser reparados, mientras el 40% restante deben ser reemplazados con unidades nuevas. Si una compañía adquiere diez de estos teléfonos, ¿cuál es la probabilidad de que exactamente dos sean reemplazados en la vigencia de su garantía?

56. La Junta de Educación reporta que 2% de los dos millones de estudiantes de preparatoria que toman el examen de aptitud escolar cada año reciben un trato especial a causa de discapacidades documentadas (*Los Ángeles Times*, 16 de julio de 2002). Considere una muestra aleatoria de 25 estudiantes que recientemente presentaron el examen.

a. ¿Cuál es la probabilidad de que exactamente 1 reciba un trato especial?

b. ¿Cuál es la probabilidad de que por lo menos 1 reciba un trato especial?

c. ¿Cuál es la probabilidad de que por lo menos 2 reciban un trato especial?

d. ¿Cuál es la probabilidad de que el número entre los 25 que recibieron un trato especial esté dentro de 2 desviaciones estándar del número que esperaría reciba un trato especial?

e. Suponga que a un estudiante que no recibe un trato especial se le permiten 3 horas para el examen, mientras que a un estudiante que recibió un trato especial se le permiten 4.5 horas. ¿Qué tiempo promedio piensa que le sería permitido a los 25 estudiantes seleccionados?

57. Suponga que 90% de todas las baterías de cierto proveedor tienen voltajes aceptables. Un tipo de linterna requiere que las dos baterías sean tipo D y funcionará sólo si sus dos baterías tienen voltajes aceptables. Entre diez linternas seleccionadas al azar, ¿cuál es la probabilidad de que por lo menos nueve funcionarán? ¿Qué suposiciones hizo para responder la pregunta planteada?

58. Un distribuidor recibe un lote muy grande de componentes. El lote sólo puede ser caracterizado como aceptable si la proporción de componentes defectuosos es cuando mucho de .10. El distribuidor decide seleccionar 10 componentes al azar y aceptar el lote sólo si el número de componentes defectuosos presentes en la muestra es cuando mucho de 2.

a. ¿Cuál es la probabilidad de que el lote será aceptado cuando la proporción real de componentes defectuosos es de .01?, ¿.05? ¿.10? ¿.20? ¿.25?

b. Sea p la proporción real de componentes defectuosos presentes en el lote. Una gráfica de P(se acepta el lote) en función de p, con p sobre el eje horizontal y P(se acepta el lote) sobre el eje vertical, se llama *curva característica de operación* del plan de muestreo de aceptación. Use los resultados del inciso (a) para trazar esta curva con $0 \le p \le 1$.

c. Repita los incisos (a) y (b) con "1" reemplazando a "2" en el plan de muestreo de aceptación.

d. Repita los incisos (a) y (b) con "15" reemplazando a "10" en el plan de muestreo de aceptación.

e. ¿Cuál de estos planes de muestreo, el del inciso (a), (c) o (d) parece más satisfactorio y por qué?

59. Un reglamento que requiere que se instale un detector de humo en todas las casas previamente construidas ha estado en vigor en una ciudad particular durante 1 año. Al departamento de bomberos le preocupa que muchas casas permanezcan sin detectores. Sea p = la proporción verdadera de las casas que tienen detectores y suponga que se inspecciona una muestra aleatoria de 25 casas. Si ésta indica marcadamente que menos de 80% de todas las casas tienen un detector, el departamento de bomberos lanzará una campaña para la puesta en ejecución de un programa de inspección obligatorio. Debido a lo caro del programa, el departamento prefiere no requerir tales inspecciones a menos que una evidencia muestral indique que se requieren. Sea X el número de casas con detectores entre las 25 muestreadas. Considere rechazar el requerimiento de que $p \ge .8$ si $x \le 15$.

a. ¿Cuál es la probabilidad de que el requerimiento sea rechazado cuando el valor real de p es .8?

b. ¿Cuál es la probabilidad de no rechazar el requerimiento cuando $p = .7$? ¿Cuando $p = .6$?

c. ¿Cómo cambian las "probabilidades de error" de los incisos (a) y (b) si el valor 15 en la regla de decisión es reemplazado por 14?

* "Entre a y b, inclusive" equivale a $(a \le X \le b)$.

60. Un puente de peaje cobra $ 1.00 para los vehículos de pasajeros y $ 2.50 para los demás vehículos. Supongamos que durante las horas del día, el 60% de todos los vehículos son de pasajeros. Si 25 vehículos cruzan el puente durante un periodo determinado durante el día, ¿cuál es el resultado de los ingresos por peaje previstos? [*Sugerencia*: Sea X = el número de vehículos de pasajeros, entonces, los ingresos por peaje $h(X)$ son una función lineal de X.]

61. Un estudiante que está tratando de escribir un ensayo para un curso tiene la opción de dos temas, A y B. Si selecciona el tema A, el estudiante pedirá dos libros mediante préstamo interbiblioteca, mientras que si selecciona el tema B, el estudiante pedirá cuatro libros. El estudiante cree que un buen ensayo necesita recibir y utilizar por lo menos la mitad de los libros pedidos para uno u otro tema seleccionado. Si la probabilidad de que un libro pedido mediante préstamo interbiblioteca llegue a tiempo es de .9 y los libros llegan independientemente uno de otro, ¿qué tema deberá seleccionar el estudiante para incrementar al máximo la probabilidad de escribir un buen ensayo? ¿Qué pasa si la probabilidad de que lleguen los libros es de sólo .5 en lugar de .9?

62. **a.** Con n fijo, ¿hay valores de p $(0 \le p \le 1)$ para los cuales $V(X) = 0$? Explique por qué esto es así.
b. ¿Con qué valor de p se incrementa al máximo $V(X)$? [*Sugerencia*: grafique $V(X)$ en función de p o bien saque una derivada.]

63. **a.** Demuestre que $b(x; n, 1 - p) = b(n - x; n, p)$.
b. Demuestre que $B(x; n, 1 - p) = 1 - B(n - x - 1; n, p)$. [*Sugerencia*: cuando mucho el número x de los S equivale a por lo menos $(n - x)$ de las F.]
c. ¿Qué implican los incisos (a) y (b) sobre la necesidad de incluir valores de p más grandes que .5 en la tabla A.1 del apéndice?

64. Demuestre que $E(X) = np$ cuando X es una variable aleatoria binomial. [*Sugerencia*: primero exprese $E(X)$ como una suma con límite inferior $x = 1$. Luego saque a np como factor, sea $y = x - 1$ de modo que la suma sea de $y = 0$ a $y = n - 1$ y demuestre que la suma es igual a 1.]

65. Los clientes en una gasolinera pagan con tarjeta de crédito (A), tarjeta de débito (B) o efectivo (C). Suponga que clientes sucesivos toman decisiones independientes con $P(A) = .5$, $P(B) = .2$ y $P(C) = .3$.
a. Entre los siguientes 100 clientes, ¿cuáles son la media y la varianza del número que paga con tarjeta de débito? Explique su razonamiento.
b. Conteste el inciso (a) para el número entre los 100 que no pagan con efectivo.

66. Una limusina de aeropuerto puede transportar hasta cuatro pasajeros en cualquier viaje. La compañía aceptará un máximo de seis reservaciones para un viaje y un pasajero debe tener una reservación. Según registros previos, 20% de los que reservan no se presentan para el viaje. Responda las siguientes preguntas, suponiendo independencia en los casos en que sea apropiado.
a. Si se hacen seis reservaciones, ¿cuál es la probabilidad de que por lo menos un individuo con reservación no pueda ser acomodado en el viaje?
b. Si se hacen seis reservaciones, ¿cuál es el número esperado de lugares disponibles cuando la limusina parte?
c. Suponga que la distribución de probabilidad del número de reservaciones hechas se da en la tabla adjunta.

Número de reservaciones	3	4	5	6
Probabilidad	.1	.2	.3	.4

Sea X el número de pasajeros en un viaje seleccionado al azar. Obtenga la función de masa de probabilidad de X.

67. Remítase a la desigualdad de Chebyshev dada en el ejercicio 44. Calcule $P(|X - \mu| \ge k\sigma)$ con $k = 2$ y $k = 3$ cuando $X \sim \text{Bin}(20, .5)$ y compare con el límite superior correspondiente. Repita para $X \sim \text{Bin}(20, .75)$.

3.5 Distribuciones hipergeométrica y binomial negativa

Las distribuciones hipergeométrica y binomial negativa están relacionadas con la distribución binomial. La distribución binomial es el modelo de probabilidad aproximada de muestreo sin reemplazo de una población dicotómica finita (S–F). Si el tamaño n de la muestra es pequeño con respecto al tamaño N de la población, la distribución hipergeométrica es el modelo de probabilidad exacta del número de éxitos (S) en la muestra. La variable aleatoria binomial X es el número de los S cuando el número n de ensayos es fijo, mientras la distribución binomial surge de fijar el número de éxitos deseados y de permitir que el número de ensayos sea aleatorio.

Distribución hipergeométrica

Las suposiciones que conducen a la distribución hipergeométrica son las siguientes:

1. La población o conjunto que se va a muestrear se compone de N individuos, objetos o elementos (una población *finita*).

2. Cada individuo puede ser caracterizado como éxito (S) o falla (F) y hay M éxitos en la población.
3. Se selecciona una muestra de n individuos sin reemplazo de tal modo que cada subconjunto de tamaño n tenga la misma probabilidad de ser seleccionado.

La variable aleatoria de interés es X = el número de las S en la muestra. La distribución de probabilidad de X depende de los parámetros n, M y N, así que se desea obtener $P(X = x) = h(x; n, M, N)$.

Ejemplo 3.35 Durante un periodo particular una oficina de tecnología de la información de una universidad recibió 20 solicitudes de servicio por problemas con impresoras, de las cuales 8 eran impresoras láser y 12 eran modelos de inyección de tinta. Se tiene que seleccionar una muestra de 5 de estas solicitudes de servicio para incluirla en una encuesta sobre satisfacción del cliente. Suponga que las 5 son seleccionadas completamente al azar, de modo que cualquier subconjunto de tamaño 5 tenga la misma probabilidad de ser seleccionado como cualquier otro subconjunto. ¿Cuál es entonces la probabilidad de que exactamente $x (x = 0, 1, 2, 3, 4 \text{ o } 5)$ de las solicitudes de servicio seleccionadas sean para impresoras de inyección de tinta?

En este caso, el tamaño de la población es $N = 20$, el tamaño de la muestra es $n = 5$ y el número de éxitos (inyección de tinta = S) y las fallas (F) en la población son $M = 12$ y $N - M = 8$, respectivamente. Considérese el valor $x = 2$. Como todos los resultados (cada uno de los cuales consta de 5 solicitudes particulares) son igualmente probables,

$$P(X = 2) = h(2; 5, 12, 20) = \frac{\text{número de resultados con } X = 2}{\text{número de resultados posibles}}$$

El número de resultados posibles en el experimento es el número de formas de seleccionar 5 de los 20 objetos sin importar el orden, es decir, $\binom{20}{5}$. Para contar el número de resultados con $X = 2$, obsérvese que existen $\binom{12}{2}$ formas de seleccionar 2 de las solicitudes para impresoras de inyección de tinta, y por cada forma existen $\binom{8}{3}$ formas de seleccionar las 3 solicitudes para impresoras láser a fin de completar la muestra. La regla de producto del capítulo 2 da entonces $\binom{12}{2}\binom{8}{3}$ como el número de resultados con $X = 2$, por lo tanto

$$h(2; 5, 12, 20) = \frac{\binom{12}{2}\binom{8}{3}}{\binom{20}{5}} = \frac{77}{323} = .238$$ ■

En general, si el tamaño de la muestra n es más pequeño que el número de éxitos en la población (M), entonces el valor de X más grande posible es n. Sin embargo, si $M < n$ (p. ej., un tamaño de muestra de 25 y sólo hay 15 éxitos en la población), entonces X puede ser cuando mucho M. Asimismo, siempre que el número de fallas en la población $(N - M)$ sobrepase el tamaño de la muestra, el valor más pequeño posible de X es 0 (puesto que todos los individuos muestreados podrían entonces ser fallas). Sin embargo, si $N - M < n$, el valor más pequeño posible de X es $n - (N - M)$. Por lo tanto, los posibles valores de X satisfacen la restricción máx $(0, n - (N - M)) \le x \le$ mín (n, M). Un argumento paralelo al del ejemplo previo da la función de masa de probabilidad de X.

PROPOSICIÓN

Si X es el número de éxitos (S) en una muestra completamente aleatoria de tamaño n extraída de la población compuesta de M éxitos y $(N - M)$ fallas, entonces la distribución de probabilidad de X, llamada **distribución hipergeométrica**, es

$$P(X = x) = h(x; n, M, N) = \frac{\binom{M}{x}\binom{N - M}{n - x}}{\binom{N}{n}} \qquad (3.15)$$

con x, un entero, que satisface máx $(0, n - N + M) \le x \le$ mín (n, M).

En el ejemplo 3.35, $n = 5$, $M = 12$ y $N = 20$, por lo tanto $h(x; 5, 12, 20)$ con $x = 0, 1, 2, 3, 4, 5$ se obtiene sustituyendo estos números en la ecuación (3.15).

Ejemplo 3.36 Se capturaron, etiquetaron y liberaron cinco individuos de una población de animales que se piensa están al borde la extinción en una región para que se mezclen con la población. Después de haber tenido la oportunidad de mezclarse, se selecciona una muestra aleatoria de 10 de estos animales. Sea $X =$ el número de animales etiquetados en la segunda muestra. Si en realidad hay 25 animales de este tipo en la región, ¿cuál es la probabilidad de que (a) $X = 2$? (b) $X \le 2$?

Los valores de los parámetros son $n = 10$, $M = 5$ (5 animales etiquetados en la población) y $N = 25$, por lo tanto

$$h(x; 10, 5, 25) = \frac{\binom{5}{x}\binom{20}{10-x}}{\binom{25}{10}} \quad x = 0, 1, 2, 3, 4, 5$$

Para el inciso (a),

$$P(X = 2) = h(2; 10, 5, 25) = \frac{\binom{5}{2}\binom{20}{8}}{\binom{25}{10}} = .385$$

Para el inciso (b),

$$P(X \le 2) = P(X = 0, 1 \text{ o } 2) = \sum_{x=0}^{2} h(x; 10, 5, 25)$$
$$= .057 + .257 + .385 = .699$$ ■

Varios paquetes de software estadístico generan fácilmente probabilidades hipergeométricas (tabular es engorroso, a causa de los tres parámetros).

Como en el caso binomial, existen expresiones simples para $E(X)$ y $V(X)$ para variables aleatorias hipergeométricas.

PROPOSICIÓN

La media y la varianza de la variable aleatoria hipergeométrica X cuya función de masa de probabilidad es $h(x; n, M, N)$ son

$$E(X) = n \cdot \frac{M}{N} \qquad V(X) = \left(\frac{N-n}{N-1}\right) \cdot n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right)$$

La razón M/N es la proporción de éxitos en la población. Si se reemplaza M/N por p en $E(X)$ y $V(X)$, se obtiene

$$E(X) = np$$
$$V(X) = \left(\frac{N-n}{N-1}\right) \cdot np(1-p) \tag{3.16}$$

La expresión (3.16) muestra que las medias de las variables aleatorias binomiales e hipergeométricas son iguales, en tanto que las varianzas de las dos variables aleatorias difieren por el factor $(N - n)/(N - 1)$, a menudo llamado **factor de corrección de población finita.** Este factor es menor que 1, así que la variable hipergeométrica tiene una varianza

más pequeña que la variable aleatoria binomial. El factor de corrección puede escribirse como $(1 - n/N)/(1 - 1/N)$, el cual es aproximadamente 1 cuando n es pequeño con respecto a N.

Ejemplo 3.37 (Continuación del ejemplo 3.36)

En el ejemplo de etiquetación de animales, $n = 10$, $M = 5$ y $N = 25$, por lo tanto $p = \frac{5}{25} = .2$ y

$$E(X) = 10(.2) = 2$$

$$V(X) = \frac{15}{24}(10)(.2)(.8) = (.625)(1.6) = 1$$

Si el muestreo se realizó con reemplazo, $V(X) = 1.6$.

Suponga que en realidad no se conoce el tamaño de la población N, así que se observa el valor x y se desea estimar N. Es razonable igualar la proporción muestral observada de éxitos, x/n, con la proporción de la población, M/N, que da la estimación

$$\hat{N} = \frac{M \cdot n}{x}$$

Si $M = 100$, $n = 40$ y $x = 16$, entonces $\hat{N} = 250$. ■

La regla general empírica dada en la sección 3.4 plantea que si el muestreo se realizó sin reemplazo pero n/N era cuando mucho de .05, entonces la distribución binomial podría ser utilizada para calcular probabilidades aproximadas que implican el número de éxitos en la muestra. Un enunciado más preciso es el siguiente: permita que el tamaño de la población N y el número de M éxitos presentes en la población se hagan más grandes a medida que la razón M/N tiende a p. Entonces $h(x; n, M, N)$ tiende a $b(x; n, p)$; así que con n/N pequeña, las dos son aproximadamente iguales siempre y cuando p no esté muy cerca de 0 o 1. Éste es la razón de ser de la regla empírica.

Distribución binomial negativa

La variable aleatoria binomial y la distribución binomial negativa se basan en un experimento que satisface las siguientes condiciones:

1. El experimento consiste en una secuencia de ensayos independientes.
2. Cada ensayo puede dar por resultado un éxito (S) o una falla (F).
3. La probabilidad de éxito es constante de un ensayo a otro, por lo tanto $P(S \text{ en el ensayo } i) = p$ con $i = 1, 2, 3, \ldots$
4. El experimento continúa (se realizan ensayos) hasta que un total de r éxitos hayan sido observados, donde r es un entero positivo especificado.

La variable aleatoria de interés es $X =$ el número de fallas que preceden al éxito r-ésimo; X se llama **variable aleatoria binomial negativa** porque, en contraste con la variable aleatoria binomial, el número de éxitos es fijo y el número de ensayos es aleatorio.

Posibles valores de X son 0, 1, 2, . . . Sea $nb(x; r, p)$ la función de masa de probabilidad de X. Considere $nb(7, 3, p) = P(X = 7)$, la probabilidad de que ocurran exactamente $7F$ antes de la 3ª S. Para que esto suceda, el décimo ensayo debe ser una S y debe haber exactamente 2 S entre los 9 primeros ensayos. Por tanto

$$nb(7; 3, p) = \left\{\binom{9}{2} \cdot p^2(1 - p)^7\right\} \cdot p = \binom{9}{2} \cdot p^3(1 - p)^7$$

La generalización de esta línea de razonamiento da la siguiente fórmula para la función de masa de probabilidad binomial negativa.

PROPOSICIÓN

La función de masa de probabilidad de la variable aleatoria binomial negativa X con los parámetros r = número de éxitos (S) y $p = P(S)$ es

$$nb(x; r, p) = \binom{x + r - 1}{r - 1} p^r (1 - p)^x \quad x = 0, 1, 2, \ldots$$

Ejemplo 3.38 Un pediatra desea reclutar 5 parejas, cada una de las cuales espera a su primer hijo, para participar en un nuevo régimen de alumbramiento natural. Sea $p = P$(una pareja seleccionada al azar está de acuerdo en participar). Si $p = .2$, ¿cuál es la probabilidad de que 15 parejas tengan que ser entrevistadas antes de encontrar 5 que estén de acuerdo en participar? Es decir, con $S = \{$está de acuerdo en participar$\}$, ¿cuál es la probabilidad de que ocurran 10 fallas antes del quinto éxito? Sustituyendo $r = 5$, $p = .2$ y $x = 10$ en $nb(x; r, p)$ da

$$nb(10; 5, .2) = \binom{14}{4}(.2)^5(.8)^{10} = .034$$

La probabilidad de que cuando mucho se observen 10 fallas (cuando mucho con 15 parejas entrevistadas) es

$$P(X \le 10) = \sum_{x=0}^{10} nb(x; 5, .2) = (.2)^5 \sum_{x=0}^{10} \binom{x + 4}{4} (.8)^x = .164$$ ■

En algunas fuentes, la variable aleatoria binomial negativa se considera como el número de ensayos $X + r$ en lugar del número de fallas.

En el caso especial $r = 1$, la función de masa de probabilidad es

$$nb(x; 1, p) = (1 - p)^x p \quad x = 0, 1, 2, \ldots \tag{3.17}$$

En el ejemplo 3.12 se dedujo la función de masa de probabilidad para el número de ensayos necesarios para obtener el primer *éxito* (S) y allí la función de masa de probabilidad es similar a la expresión (3.17). En la literatura se hace referencia tanto a X = número de *fallas* (F) como a Y = número de ensayos ($= 1 + X$) como **variables aleatorias geométricas,** y la función de masa de probabilidad en la expresión (3.17) se llama **distribución geométrica.**

En el ejemplo 3.19 se demostró que el número esperado de ensayos hasta que aparece el primer éxito es $1/p$, así que el número esperado de *fallas* hasta que aparece el primer éxito es $(1/p) - 1 = (1 - p)/p$. Intuitivamente, se esperaría ver $r \cdot (1 - p)/p$ fallas antes del éxito r-ésimo y éste en realidad es $E(X)$. También existe una fórmula simple para $V(X)$.

PROPOSICIÓN

Si X es una variable aleatoria binomial negativa con función de masa de probabilidad $nb(x; r, p)$, entonces

$$E(X) = \frac{r(1 - p)}{p} \qquad V(X) = \frac{r(1 - p)}{p^2}$$

Por último, al expandir el coeficiente binomial enfrente de $p^r(1 - p)^x$ y haciendo alguna cancelación, se ve que $nb(x; r, p)$ está bien definido incluso cuando r no es un entero. Se ha encontrado que la *distribución binomial negativa generalizada* para ajustar los datos observados verdaderamente bien en una amplia variedad de aplicaciones.

EJERCICIOS Sección 3.5 (68–78)

68. Una tienda de electrónica ha recibido un envío de 20 radios de mesa que tienen conexiones para el iPod o iPhone. Doce de ellos tienen dos ranuras (para que puedan acomodar a los dos dispositivos) y los otros ocho tienen una sola ranura. Supongamos que seis de los 20 radios son seleccionados al azar para ser almacenados en un estante donde son exhibidos y los restantes se colocan en un almacén. Sea X = el número de los radios almacenados en el estante de exhibición que tienen dos ranuras.
- **a.** ¿Qué clase de distribución tiene X (nombre y valores de todos los parámetros)?
- **b.** Calcule $P(X = 2)$, $P(X \le 2)$ y $P(X \ge 2)$.
- **c.** Calcule el valor medio y la desviación estándar de X.

69. Cada uno de 12 refrigeradores de un tipo ha sido regresado a un distribuidor debido a un ruido agudo audible producido por oscilación cuando el refrigerador está funcionando. Suponga que 7 de estos refrigeradores tienen un compresor defectuoso y que los otros 5 tienen problemas menos serios. Si los refrigeradores se examinan en orden aleatorio, sea X el número entre los primeros 6 examinados que tienen un compresor defectuoso. Calcule lo siguiente:
- **a.** $P(X = 5)$
- **b.** $P(X \le 4)$
- **c.** La probabilidad de que X exceda su valor medio por más de 1 desviación estándar.
- **d.** Considere un gran envío de 400 refrigeradores, 40 de los cuales tienen compresores defectuosos. Si X es el número entre 15 refrigeradores seleccionados al azar que tienen compresores defectuosos, describa una forma menos tediosa de calcular (por lo menos de forma aproximada) $P(X \le 5)$ que utilizar la función de masa de probabilidad hipergeométrica.

70. Un instructor que impartió dos secciones de estadística para ingeniería el semestre pasado, la primera con 20 estudiantes y la segunda con 30, decidió asignar un proyecto semestral. Una vez que todos los proyectos le fueron entregados, el instructor los ordenó al azar antes de calificarlos. Considere los primeros 15 proyectos calificados.
- **a.** ¿Cuál es la probabilidad de que exactamente 10 de éstos sean de la segunda sección?
- **b.** ¿Cuál es la probabilidad de que por lo menos 10 de éstos sean de la segunda sección?
- **c.** ¿Cuál es la probabilidad de que por lo menos 10 de éstos sean de la misma sección?
- **d.** ¿Cuáles son el valor medio y la desviación estándar del número entre estos 15 que son de la segunda sección?
- **e.** ¿Cuáles son el valor medio y la desviación estándar del número de proyectos que no están entre estos primeros 15 que son de la segunda sección?

71. Un geólogo recolectó 10 especímenes de roca basáltica y 10 de granito. Él le pide a su ayudante de laboratorio que seleccione al azar 15 de los especímenes para analizarlos.
- **a.** ¿Cuál es la función de masa de probabilidad del número de especímenes de granito seleccionados para su análisis?
- **b.** ¿Cuál es la probabilidad de que todos los especímenes de uno de los dos tipos de roca sean seleccionados para su análisis?
- **c.** ¿Cuál es la probabilidad de que el número de especímenes de granito seleccionados para analizarlos esté dentro de 1 desviación estándar de su valor medio?

72. Un director de personal que va a entrevistar a 11 ingenieros para cuatro vacantes de trabajo ha programado seis entrevistas para el primer día y cinco para el segundo. Suponga que los candidatos son entrevistados en orden aleatorio.
- **a.** ¿Cuál es la probabilidad de que x de los cuatro mejores candidatos sean entrevistados el primer día?
- **b.** ¿Cuántos de los mejores cuatro candidatos se espera que puedan ser entrevistados el primer día?

73. Veinte parejas de individuos que participan en un torneo de bridge han sido sembrados del 1, . . . , 20. En esta primera parte del torneo, los 20 son divididos al azar en 10 parejas este–oeste y 10 parejas norte–sur.
- **a.** ¿Cuál es la probabilidad de que x de las 10 mejores parejas terminen jugando este–oeste?
- **b.** ¿Cuál es la probabilidad de que las cinco mejores parejas terminen jugando en la misma dirección?
- **c.** Si existen $2n$ parejas, ¿cuál es la función de masa de probabilidad de X = el número entre las mejores n parejas que terminan jugando este–oeste? ¿Cuáles son $E(X)$ y $V(X)$?

74. Una alerta contra smog de segunda etapa ha sido emitida en un área del condado de Los Ángeles en la cual hay 50 firmas industriales. Un inspector visitará 10 firmas seleccionadas al azar para ver si no han violado los reglamentos.
- **a.** Si 15 de las firmas sí están violando por lo menos un reglamento, ¿cuál es la función de masa de probabilidad del número de firmas visitadas por el inspector que violan por lo menos un reglamento?
- **b.** Si existen 500 firmas en el área, 150 de las cuales violan algún reglamento, represente de forma aproximada la función de masa de probabilidad del inciso (a) con una función de masa de probabilidad más simple.
- **c.** Con X = el número entre las 10 visitadas que violan algún reglamento, calcule $E(X)$ y $V(X)$ ambas para la función de masa de probabilidad exacta y la función de masa de probabilidad aproximada del inciso (b).

75. Suponga que $p = P(\text{nacimiento de un varón}) = .5$. Una pareja desea tener exactamente dos niñas en su familia. Tendrán hijos hasta que esta condición se satisfaga.
- **a.** ¿Cuál es la probabilidad de que la familia tenga x varones?
- **b.** ¿Cuál es la probabilidad de que la familia tenga cuatro hijos?
- **c.** ¿Cuál es la probabilidad de que la familia tenga cuando mucho cuatro hijos?
- **d.** ¿Cuántos varones cree que tendría esta familia? ¿Cuántos hijos esperaría que tenga esta familia?

76. Una familia decide tener hijos hasta que tengan tres del mismo sexo. Suponiendo $P(B) = P(G) = .5$, ¿cuál es la función de masa de probabilidad de X = el número de hijos en la familia?

77. Tres hermanos y sus esposas deciden tener hijos hasta que cada familia tenga dos niñas. ¿Cuál es la función de masa de probabilidad de X = el número total de varones procreados por los hermanos? ¿Cuál es $E(X)$ y cómo se compara con el número esperado de varones procreados por cada hermano?

78. De acuerdo con el artículo "Characterizing the Severity and Risk of Drought in the Poudre River, Colorado" (*J. of Water Res. Planning and Mgmnt.*, 2005: 383–393), la longitud de la sequía Y es el número de intervalos de tiempo consecutivos en los que el suministro de agua se mantiene por debajo de un valor crítico y_0 (un déficit), precedido y seguido por periodos en los que el suministro supera este valor crítico (un excedente). El documento citado propone una distribución geométrica con $p = .409$ para esta variable aleatoria.

a. ¿Cuál es la probabilidad de que una sequía dure exactamente 3 intervalos? ¿A lo más 3 intervalos?

b. ¿Cuál es la probabilidad de que la duración de una sequía exceda su valor medio por al menos una desviación estándar?

3.6 Distribución de probabilidad de Poisson

Las distribuciones binomial, hipergeométrica y binomial negativa se dedujeron partiendo de un experimento compuesto de ensayos o sorteos y aplicando las leyes de probabilidad a varios resultados del experimento. No existe un experimento simple en el cual esté basada la distribución de Poisson, aun cuando en breve se describirá cómo puede ser obtenida mediante ciertas operaciones restrictivas.

DEFINICIÓN

Se dice que una variable aleatoria discreta X tiene una **distribución de Poisson** con parámetro μ ($\mu > 0$) si la función de masa de probabilidad de X es

$$p(x; \mu) = \frac{e^{-\mu} \cdot \mu^x}{x!} \qquad x = 0, 1, 2, 3, \ldots$$

No es casualidad que se esté usando el símbolo μ para el parámetro de Poisson, en breve se verá que μ es en realidad el valor esperado de X. La letra e en la función de masa de probabilidad representa la base del sistema de logaritmos naturales, su valor numérico es aproximadamente 2.71828. A diferencia de las distribuciones binomial e hipergeométrica, la distribución de probabilidad de Poisson se extiende a *todos* los números enteros no negativos, un número infinito de posibilidades.

No es evidente por inspección que $p(x; \mu)$ especifique una función de masa de probabilidad legítima, por no hablar de que esta distribución es útil. En primer lugar, $p(x; \mu) > 0$ para cada valor x posible debido a la exigencia de que $\mu > 0$. El hecho de que $\sum p(x; \mu) = 1$ es una consecuencia de la expansión en series de Maclaurin de e^{μ} (consulte su libro de cálculo para este resultado):

$$e^{\mu} = 1 + \mu + \frac{\mu^2}{2!} + \frac{\mu^3}{3!} + \cdots = \sum_{x=0}^{\infty} \frac{\mu^x}{x!} \tag{3.18}$$

Si los dos términos extremos de la expresión (3.18) se multiplican por $e^{-\mu}$ y luego esta cantidad se coloca adentro de la suma en el lado derecho, el resultado es

$$1 = \sum_{x=0}^{\infty} \frac{e^{-\mu} \cdot \mu^x}{x!}$$

Ejemplo 3.39 Sea X el número de criaturas de un tipo particular capturadas en una trampa durante un lapso de tiempo dado. Suponga que X tiene una distribución de Poisson con $\mu = 4.5$, así que en promedio las trampas contendrán 4.5 criaturas. [El artículo "Dispersal Dynamics of

the Bivalve *Gemma gemma* in a Patchy Environment" (*Ecological Monographs*, 1995: 1–20) sugiere este modelo: el molusco bivalvo *Gemma gemma* es una pequeña almeja.] La probabilidad de que una trampa contenga exactamente cinco criaturas es

$$P(X = 5) = \frac{e^{-4.5}(4.5)^5}{5!} = .1708$$

La probabilidad de que una trampa contenga cuando mucho cinco criaturas es

$$P(X \le 5) = \sum_{x=0}^{5} \frac{e^{-4.5}(4.5)^x}{x!} = e^{-4.5}\left[1 + 4.5 + \frac{(4.5)^2}{2!} + \cdots + \frac{(4.5)^5}{5!}\right] = .7029$$ ■

La distribución de Poisson como límite

La siguiente proposición suministra la razón de ser en el uso de la distribución de Poisson en muchas situaciones.

PROPOSICIÓN

Suponga que en la función de masa de probabilidad binomial $b(x; n, p)$, $n \to \infty$ y $p \to 0$ de tal modo que np tienda a un valor $\mu > 0$. Entonces $b(x; n, p) \to p(x; \mu)$.

De acuerdo con esta proposición, *en cualquier experimento binomial en el cual n es grande y p es pequeña*, $b(x; n, p) \approx p(x; \mu)$, donde $\mu = np$. Como regla empírica, esta aproximación puede ser aplicada con seguridad si $n > 50$ y $np < 5$.

Ejemplo 3.40 Si un editor de libros no técnicos hace todo lo posible por que sus libros estén libres de errores tipográficos, de modo que la probabilidad de que cualquier página dada contenga por lo menos uno de esos errores es de .005 y los errores son independientes de una página a otra, ¿cuál es la probabilidad de que una de sus novelas de 400 páginas contenga exactamente una página con errores? ¿Cuándo mucho tres páginas con errores?

Con *S* denotando una página que contiene por lo menos un error y *F* una página libre de errores, el número *X* de páginas que contienen por lo menos un error es una variable aleatoria binomial con $n = 400$ y $p = .005$, así que $np = 2$. Se desea

$$P(X = 1) = b(1; 400, .005) \approx p(1; 2) = \frac{e^{-2}(2)^1}{1!} = .270671$$

El valor binomial es $b(1; 400, .005) = .270669$, así que la aproximación es muy buena.

Asimismo,

$$\begin{aligned} P(X \le 3) &\approx \sum_{x=0}^{3} p(x, 2) = \sum_{x=0}^{3} e^{-2}\frac{2^x}{x!} \\ &= .135335 + .270671 + .270671 + .180447 \\ &= .8571 \end{aligned}$$

y éste de nuevo se aproxima bastante al valor binomial $P(X \le 3) = .8576$. ■

La tabla 3.2 muestra la distribución de Poisson con $\mu = 3$ junto con tres distribuciones binomiales con $np = 3$ y la figura 3.8 (generada por S-Plus) ilustra una gráfica de la distribución de Poisson junto con las dos primeras distribuciones binomiales. La aproximación es de uso limitado con $n = 30$, pero desde luego la precisión es mejor con $n = 100$ y mucho mejor con $n = 300$.

Tabla 3.2 Comparación de la distribución de Poisson y tres distribuciones binomiales

x	$n = 30, p = .1$	$n = 100, p = .03$	$n = 300, p = .01$	Poisson, $\mu = 3$
0	0.042391	0.047553	0.049041	0.049787
1	0.141304	0.147070	0.148609	0.149361
2	0.227656	0.225153	0.224414	0.224042
3	0.236088	0.227474	0.225170	0.224042
4	0.177066	0.170606	0.168877	0.168031
5	0.102305	0.101308	0.100985	0.100819
6	0.047363	0.049610	0.050153	0.050409
7	0.018043	0.020604	0.021277	0.021604
8	0.005764	0.007408	0.007871	0.008102
9	0.001565	0.002342	0.002580	0.002701
10	0.000365	0.000659	0.000758	0.000810

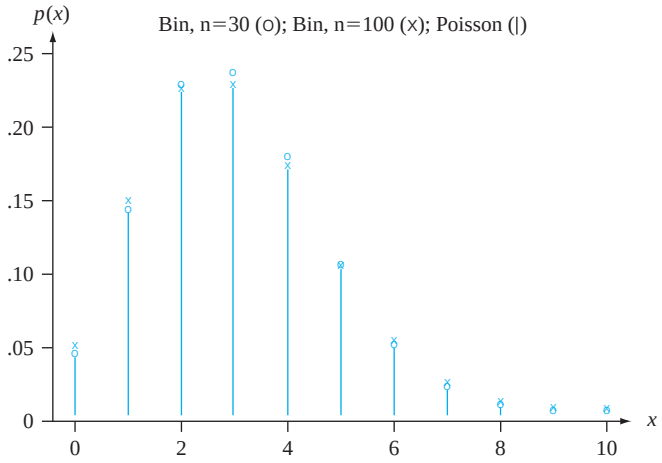


Figura 3.8 Comparación entre una distribución de Poisson y dos distribuciones binomiales

La tabla A.2 del apéndice muestra la función de distribución acumulativa $F(x; \mu)$ para $\mu = .1, .2, \ldots, 1, 2, \ldots, 10, 15$ y 20. Por ejemplo, si $\mu = 2$, entonces $P(X \leq 3) = F(3; 2) = .857$ como en el ejemplo 3.40, en tanto que $P(X = 3) = F(3; 2) - F(2; 2) = .180$. Alternativamente, muchos paquetes de computadora estadísticos generarán $p(x; \mu)$ y $F(x; \mu)$ al solicitarlo.

Media y varianza de X

Como $b(x; n, p) \to p(x; \mu)$ a medida que $n \to \infty$, $p \to 0$, $np \to \mu$, la media y la varianza de una variable binomial deberán aproximarse a las de una variable de Poisson. Estos límites son $np \to \mu$ y $np(1 - p) \to \mu$.

PROPOSICIÓN

Si X tiene una distribución de Poisson con parámetro μ, entonces $E(X) = V(X) = \mu$.

Estos resultados también pueden obtenerse directamente de la definición de media y varianza.

Ejemplo 3.41 (Continuación del ejemplo 3.39)

Tanto el número esperado de criaturas atrapadas como la varianza de éste son iguales a 4.5, y $\sigma_X = \sqrt{\mu} = \sqrt{4.5} = 2.12$. ■

Proceso de Poisson

Una aplicación muy importante de la distribución de Poisson surge en conexión con la ocurrencia de eventos de algún tipo en el transcurso del tiempo. Eventos de interés podrían ser visitas a un sitio web particular, pulsos de alguna clase registrados por un contador, mensajes de correo electrónico enviados a una dirección particular, accidentes en una instalación industrial o lluvias de rayos cósmicos observados por astrónomos en un observatorio particular. Se hace la siguiente suposición sobre la forma en que los eventos de interés ocurren:

1. Existe un parámetro $\alpha > 0$ tal que durante cualquier intervalo de tiempo corto Δt, la probabilidad de que ocurra exactamente un evento es $\alpha \cdot \Delta t + o(\Delta t)$*
2. La probabilidad de que ocurra más de un evento durante Δt es $o(\Delta t)$ [la que junto con la suposición 1, implica que la probabilidad de ningún evento durante Δt es $1 - \alpha \cdot \Delta t - o(\Delta t)$.
3. El número de eventos ocurridos durante este intervalo de tiempo Δt es independiente del número ocurrido antes de este intervalo de tiempo.

Informalmente, la suposición 1 dice que durante un corto intervalo de tiempo, la probabilidad de que ocurra un solo evento es aproximadamente proporcional a la duración del intervalo de tiempo, donde α es la constante de proporcionalidad. Ahora sea $P_k(t)$ la probabilidad de que k eventos serán observados durante cualquier intervalo de tiempo particular de duración t.

PROPOSICIÓN

$P_k(t) = e^{-\alpha t} \cdot (\alpha t)^k/k!$, de modo que el número de eventos durante un intervalo de tiempo de duración t es una variable aleatoria de Poisson con parámetro $\mu = \alpha t$. El número esperado de eventos durante cualquier intervalo de tiempo es entonces αt, así que el número esperado durante un intervalo de tiempo unitario es α.

La ocurrencia de eventos en el transcurso del tiempo como se describió se llama *proceso de Poisson*; el parámetro α especifica *la rapidez* del proceso.

Ejemplo 3.42 Suponga que llegan pulsos a un contador a un ritmo promedio de seis por minuto, así que $\alpha = 6$. Para determinar la probabilidad de que en un intervalo de .5 minuto se reciba por lo menos un pulso, obsérvese que el número de pulsos en ese intervalo tiene una distribución de Poisson con parámetro $\alpha t = 6(.5) = 3$ (se utiliza .5 minuto porque α está expresada como rapidez por minuto). Entonces con $X =$ el número de pulsos recibidos en el intervalo de 30 segundos,

$$P(1 \le X) = 1 - P(X = 0) = 1 - \frac{e^{-3}(3)^0}{0!} = .950$$ ■

En lugar de observar eventos en el transcurso del tiempo, considere observar eventos de algún tipo que ocurren en una región de dos o tres dimensiones. Por ejemplo, se podría seleccionar un mapa de una región R de un bosque, ir a dicha región y contar el número de árboles. Cada árbol representaría un evento que ocurre en un punto particular del espacio. Conforme a suposiciones similares a 1–3, se puede demostrar que el número de eventos que ocurren en una región R tiene una distribución de Poisson con parámetro $\alpha \cdot a(R)$, donde $a(R)$ es el área de R. La cantidad α es el número esperado de eventos por unidad de área o volumen.

* Una cantidad es $o(\Delta t)$ (léase "o pequeña de delta t") si, a medida que t tiende a cero, también lo hace $o(\Delta t)/\Delta t$. Es decir, $o(\Delta t)$ es incluso más insignificante (tiende a 0 más rápido) que Δt mismo. La cantidad $(\Delta t)^2$ tiene esta propiedad, pero $\text{sen}(\Delta t)$ no.

EJERCICIOS Sección 3.6 (79–93)

79. Sea X el número de imperfecciones superficiales de una caldera seleccionada al azar de un tipo que tiene una distribución de Poisson con parámetro $\mu = 5$. Use la tabla A.2 del apéndice para calcular las siguientes probabilidades
a. $P(X \le 8)$ **b.** $P(X = 8)$ **c.** $P(9 \le X)$
d. $P(5 \le X \le 8)$ **e.** $P(5 < X < 8)$

80. Sea X el número de anomalías que ocurren en el material de una región particular de un disco de turbina de gas en aviones. El artículo "Methodology for Probabilistic Life Prediction of Multiple-Anomaly Materials" (*Amer. Inst. of Aeronautics and Astronautics J.*, 2006: 787–793) propone una distribución de Poisson para X. Supongamos que $\mu = 4$.
a. Calcule $P(X \le 4)$ y $P(X < 4)$.
b. Calcule $P(4 \le X \le 8)$.
c. Calcule $P(8 \le X)$.
d. ¿Cuál es la probabilidad de que el número observado de anomalías sobrepase suvalor medio por no más de una desviación estándar?

81. Suponga que el número de conductores que viajan entre un origen y destino particulares durante un lapso de tiempo designado tiene una distribución de Poisson con parámetro $\mu = 20$ (sugerido en el artículo "Dynamic Ride Sharing: Theory and Practice", *J. of Transp. Engr.*, 1997:308–312). ¿Cuál es la probabilidad de que el número de conductores
a. será cuando mucho de 10?
b. será de más de 20?
c. será de entre 10 y 20, inclusive? Será estrictamente de entre 10 y 20?
d. estará dentro de 2 desviaciones estándar del valor medio?

82. Considere escribir en un disco de computadora y luego enviarlo a través de un certificador que cuenta el número de pulsos faltantes. Suponga que este número X tiene una distribución de Poisson con parámetro $\mu = .2$. (Sugerido en "Average Sample Number for Semi-Curtailed Sampling Using the Poisson Distribution", *J. Quality Technology*, 1983: 126–129.)
a. ¿Cuál es la probabilidad de que un disco tenga exactamente un pulso faltante?
b. ¿Cuál es la probabilidad de que un disco tenga por lo menos dos pulsos faltantes?
c. Si se seleccionan dos discos independientemente, ¿cuál es la probabilidad de que ninguno contenga un pulso faltante?

83. Un artículo en *Los Ángeles Times* (3 de diciembre de 1993) reporta que 1 de cada 200 personas porta el gen defectuoso que provoca cáncer de colon hereditario. En una muestra de 1000 individuos, ¿cuál es la distribución aproximada del número que porta este gen? Use esta distribución para calcular la probabilidad aproximada de que
a. Entre 5 y 8 (inclusive) porten el gen.
b. Por lo menos 8 porten el gen.

84. Suponga que sólo .10% de todas las computadoras de cierto tipo experimentan fallas del CPU durante el periodo de garantía. Considere una muestra de 10,000 computadoras.
a. ¿Cuáles son el valor esperado y la desviación estándar del número de computadoras en la muestra que tienen el defecto?
b. ¿Cuál es la probabilidad (aproximada) de que más de 10 computadoras muestreadas tengan el defecto?
c. ¿Cuál es la probabilidad (aproximada) de que ninguna computadora muestreada tenga el defecto?

85. Suponga que una pequeña aeronave aterriza en un aeropuerto de acuerdo con un proceso de Poisson con razón $\alpha = 8$ por hora, de modo que el número de aterrizajes durante un lapso de tiempo de t horas es una variable aleatoria de Poisson con parámetro $\mu = 8t$.
a. ¿Cuál es la probabilidad de que exactamente 6 aeronaves pequeñas aterricen durante un intervalo de 1 hora? ¿Por lo menos 6? ¿Por lo menos 10?
b. ¿Cuáles son el valor esperado y la desviación estándar del número de aeronaves pequeñas que aterrizan durante un lapso de 90 min?
c. ¿Cuál es la probabilidad de que por lo menos 20 aeronaves pequeñas aterricen durante un lapso de 2.5 horas? ¿De qué cuando mucho aterricen 10 durante este periodo?

86. El número de personas que llegan para tratamiento a una sala de urgencias puede ser modelado mediante un proceso de Poisson con parámetro de rapidez de cinco por hora.
a. ¿Cuál es la probabilidad de que ocurran exactamente cuatro arribos durante una hora particular?
b. ¿Cuál es la probabilidad de que por lo menos cuatro personas arriben durante una hora particular?
c. ¿Cuántas personas espera que arriben durante un periodo de 45 min?

87. El número de solicitudes de ayuda recibidas por un servicio de grúas es un proceso de Poisson con razón $\alpha = 4$ por hora.
a. Calcule la probabilidad de que exactamente diez solicitudes sean recibidas durante un periodo particular de 2 horas.
b. Si los operadores del servicio de grúas hacen una pausa de 30 minutos para el almuerzo, ¿cuál es la probabilidad de que no dejen de atender llamadas de ayuda?
c. ¿Cuántas llamadas esperaría durante esta pausa?

88. Al someter a prueba tarjetas de circuito, la probabilidad de que cualquier diodo particular falle es de .01. Suponga que una tarjeta de circuito contiene 200 diodos.
a. ¿Cuántos diodos esperaría que fallen y cuál es la desviación estándar del número que se espera fallen?
b. ¿Cuál es la probabilidad (aproximada) de que por lo menos cuatro diodos fallen en una tarjeta seleccionada al azar?
c. Si se envían cinco tarjetas a un cliente particular, ¿qué tan probable es que por lo menos cuatro de ellas funcionen apropiadamente? (Una tarjeta funciona apropiadamente sólo si todos sus diodos funcionan.)

89. El artículo "Reliability-Based Service-Life Assessment of Aging Concrete Structures" (*J. Structural Engr.*, 1993: 1600–1621) sugiere que un proceso de Poisson puede ser utilizado para representar la ocurrencia de cargas estructurales en el transcurso del tiempo. Suponga que el tiempo medio entre ocurrencias de cargas es de .5 al año.
a. ¿Cuántas cargas se espera que ocurran durante un periodo de 2 años?

b. ¿Cuál es la probabilidad de que ocurran más de cinco cargas durante un periodo de 2 años?

c. ¿Qué tan largo debe ser un periodo de modo que la probabilidad de que no ocurran cargas durante dicho periodo sea cuando mucho de .1?

90. Sea X que tiene una distribución de Poisson con parámetro μ. Demuestre que $E(X) = \mu$ derivada directamente de la definición de valor esperado. [*Sugerencia*: el primer término en la suma es igual a 0 y luego x puede ser eliminada. Ahora saque como factor a μ y demuestre que lo que queda suma 1.]

91. Suponga que hay árboles distribuidos en un bosque de acuerdo con un proceso de Poisson bidimensional con parámetro α, el número esperado de árboles por acre es de 80.

a. ¿Cuál es la probabilidad de que en un terreno de cuarto de acre, habrá cuando mucho 16 árboles?

b. Si el bosque abarca 85,000 acres, ¿cuál es el número esperado de árboles en el bosque?

c. Suponga que selecciona un punto en el bosque y construye un círculo de .1 milla de radio. Sea X = el número de árboles dentro de esa región circular. ¿Cuál es la función de masa de probabilidad de X? [*Sugerencia*: 1 milla cuadrada = 640 acres.]

92. A una estación de inspección de equipo vehicular llegan automóviles de acuerdo con un proceso de Poisson con razón $\alpha = 10$ por hora. Suponga que un vehículo que llega con probabilidad de .5 no tendrá violaciones de equipo.

a. ¿Cuál es la probabilidad de que exactamente diez lleguen durante la hora y que ninguno tenga violaciones?

b. Con cualquier $y \geq 10$ fija, ¿cuál es la probabilidad de que y lleguen durante la hora, diez de los cuales no tengan violaciones?

c. ¿Cuál es la probabilidad de que lleguen diez carros "sin violaciones" durante la siguiente hora? [*Sugerencia*: sume las probabilidades en el inciso (b) desde $y = 10$ hasta ∞.]

93. **a.** En un proceso de Poisson, ¿qué tiene que suceder tanto en el intervalo de tiempo $(0, t)$ como en el intervalo $(t, t + \Delta t)$ de modo que no ocurran eventos en todo el intervalo $(0, t + \Delta t)$? Use esto y las suposiciones 1–3 para escribir una relación entre $P_0(t + \Delta t)$ y $P_0(t)$.

b. Use el resultado del inciso (a) para escribir una expresión para la diferencia $P_0(t + \Delta t) - P_0(t)$. Divida entonces entre Δt y permita que $\Delta t \to 0$ para obtener una ecuación que implique $(d/dt)P_0(t)$, la derivada de $P_0(t)$ con respecto a t.

c. Verifique que $P_0(t) = e^{-\alpha t}$ satisface la ecuación del inciso (b).

d. Se puede demostrar de manera similar a los incisos (a) y (b) que los $P_k(t)$ deben satisfacer el sistema de ecuaciones diferenciales

$$\frac{d}{dt}P_k(t) = \alpha P_{k-1}(t) - \alpha P_k(t)$$
$$k = 1, 2, 3, \ldots$$

Verifique que $P_k(t) = e^{-\alpha t}(\alpha t)^k/k!$ satisface el sistema. (En realidad ésta es la única solución.)

EJERCICIOS SUPLEMENTARIOS (94–122)

94. Considere un mazo compuesto de siete cartas, marcadas 1, 2, . . . , 7. Se seleccionan al azar tres de estas cartas. Defina una variable aleatoria W como W = la suma de los números resultantes y calcule la función de masa de probabilidad de W. Calcule entonces μ y σ^2. [*Sugerencia*: considere los resultados sin orden, de modo que (1, 3, 7) y (3, 1, 7) no son resultados diferentes. Entonces existen 35 resultados y pueden ser puestos en lista. (Este tipo de variable aleatoria en realidad se presenta en conexión con una prueba de hipótesis llamada prueba de suma de renglones de Wilcoxon, en la cual hay una muestra x y una muestra y y W es la suma de los renglones de x en la muestra combinada; véase la sección 15.2.)

95. Después de barajar un mazo de 52 cartas, un tallador reparte 5. Sea X = el número de palos representados en la mano de 5 cartas.

a. Demuestre que la función de masa de probabilidad de X es

x	1	2	3	4
$p(x)$	.002	.146	.588	.264

[*Sugerencia*: $p(1) = 4P$(todas son espadas), $p(2) = 6P$(sólo espadas y corazones con por lo menos una de cada palo) y $p(4) = 4P$(2 espadas $\cap$ una de cada otro palo).]

b. Calcule μ, σ^2 y σ.

96. La variable aleatoria binomial negativa X se definió como el número de fallas (F) que preceden al éxito (S) r-ésimo. Sea Y = el número de ensayos necesarios para obtener el *éxito* (S) r-ésimo. Del mismo modo en que fue obtenida la función de masa de probabilidad de X, deduzca la función de masa de probabilidad de Y.

97. De todos los clientes que adquieren abrepuertas de cochera automáticos, 75% adquieren el modelo de transmisión por cadena. Sea X = el número entre los siguientes 15 compradores que seleccionan el modelo de transmisión por cadena.

a. ¿Cuál es la función de masa de probabilidad de X?

b. Calcule $P(X > 10)$.

c. Calcule $P(6 \leq X \leq 10)$.

d. Calcule μ y σ^2.

e. Si la tienda actualmente tiene en existencia 10 modelos de transmisión por cadena y 8 modelos de transmisión por flecha, ¿cuál es la probabilidad de que las solicitudes de estos 15 clientes puedan ser satisfechas con las existencias actuales?

98. Un amigo recientemente planeó un viaje de campamento. Tenía dos linternas, una que requería una sola batería de 6 V y otra que utilizaba dos baterías de tamaño D. Antes había empacado dos baterías de 6 V y cuatro tamaño D en su camper. Suponga que la probabilidad de que cualquier batería particular funcione es p y que las baterías funcionan o fallan indepen-

dientemente una de otra. Nuestro amigo desea llevar sólo una linterna. ¿Con qué valores de p deberá llevar la linterna de 6 V?

99. Un *sistema k de n* es uno que funcionará si y sólo si por lo menos k de los n componentes individuales en el sistema funcionan. Si los componentes individuales funcionan independientemente uno de otro, cada uno con probabilidad de .9, ¿cuál es la probabilidad de que un sistema 3 de 5 funcione?

100. Un fabricante de chips de circuitos integrados desea controlar la calidad de sus productos rechazando cualquier lote en el que la proporción de chips sea demasiado alta. Con esta finalidad, de cada lote de 10,000 chips, se seleccionarán y probarán 25. Si por lo menos 5 de éstos están defectuosos, todo el lote será rechazado.

a. ¿Cuál es la probabilidad de que un lote será rechazado si 5% de los chips en el lote están de hecho, defectuosos?

b. Responda la pregunta del inciso (a) si el porcentaje de chips defectuosos es 10%.

c. Responda la pregunta del inciso (a) si el porcentaje de chips defectuosos es 20%.

d. ¿Qué les sucedería a las probabilidades en los incisos (a)–(c) si el número de rechazo crítico se incrementara de 5 a 6?

101. De las personas que pasan a través de un detector de metales en un aeropuerto, el .5% lo activan; sea X = el número entre un grupo de 500 seleccionado al azar que activan el detector.

a. ¿Cuál es la función de masa de probabilidad (aproximada) de X?

b. Calcule $P(X = 5)$.

c. Calcule $P(5 \leq X)$.

102. Una firma consultora educativa está tratando de decidir si los estudiantes de preparatoria que nunca antes han utilizado una calculadora de mano pueden resolver cierto tipo de problema más fácilmente con una calculadora que utiliza lógica polaca inversa o una que no utiliza esta lógica. Se selecciona una muestra de 25 estudiantes y se les permite practicar con ambas calculadoras. Luego a cada estudiante se le pide que resuelva un problema con la calculadora polaca inversa y un problema similar con la otra. Sea $p = P(S)$, donde S indica que un estudiante resolvió el problema más rápido con la lógica polaca inversa que sin ella y sea X = número de éxitos.

a. Si $p = .5$, ¿cuál es $P(7 \leq X \leq 18)$?

b. Si $p = .8$, ¿cuál es $P(7 \leq X \leq 18)$?

c. Si la pretensión de que $p = .5$ tiene que ser rechazada cuando $x \leq 7$ o $x \geq 18$, ¿cuál es la probabilidad de rechazar la pretensión cuando en realidad es correcta?

d. Si la decisión de rechazar la pretensión $p = .5$ se hace como en el inciso (c), ¿cuál es la probabilidad de que la pretensión no sea rechazada cuando $p = .6$? ¿Cuándo $p = .8$?

e. ¿Qué regla de decisión escogería para rechazar la pretensión de que $p = .5$ si desea que la probabilidad en el inciso (c) sea cuando mucho de .01?

103. Considere una enfermedad cuya presencia puede ser identificada por medio de un análisis de sangre. Sea p la probabilidad de que un individuo seleccionado al azar tenga la enfermedad. Suponga que se seleccionan independientemente n individuos para analizarlos. Una forma de proceder es analizar cada una de las n muestras de sangre. Un procedimiento potencialmente más económico, de análisis en grupo, se introdujo durante la Segunda Guerra Mundial para identificar hombres sifilíticos entre los reclutas. En primer lugar, se toma una parte de cada muestra de sangre, se combinan estos especímenes y se realiza un solo análisis. Si ninguno tiene la enfermedad, el resultado será negativo y sólo se requiere un análisis. Si por lo menos un individuo está enfermo, el análisis de la muestra combinada dará un resultado positivo, en cuyo caso se realizan los análisis de los n individuos. Si $p = .1$ y $n = 3$, ¿cuál es el número esperado de análisis si se utiliza este procedimiento? ¿Cuál es el número esperado cuando $n = 5$? [El artículo "Random Multiple-Access Communication and Group Testing" (*IEEE Trans. on Commun.*, 1984: 769–774) aplicó estas ideas a un sistema de comunicación en el cual la dicotomía fue usuario ocioso/activo en lugar de enfermo/no enfermo.]

104. Sea p_1 la probabilidad de que cualquier símbolo de código particular sea erróneamente transmitido a través de un sistema de comunicación. Suponga que en diferentes símbolos, ocurren errores de manera independiente uno de otro. Suponga también que con probabilidad p_2 un símbolo erróneo es corregido al ser recibido. Sea X el número de símbolos correctos en un bloque de mensaje compuesto de n símbolos (una vez que el proceso de corrección ha terminado). ¿Cuál es la distribución de probabilidad de X?

105. El comprador de una unidad generadora de potencia requiere de c arranques consecutivos exitosos antes de aceptar la unidad. Suponga que los resultados de arranques individuales son independientes entre sí. Sea p la probabilidad de que cualquier arranque particular sea exitoso. La variable aleatoria de interés es X = el número de arranques que deben hacerse antes de la aceptación. Dé la función de masa de probabilidad de X en el caso $c = 2$. Si $p = .9$, ¿cuál es $P(X \leq 8)$? [*Sugerencia*: con $x \geq 5$, exprese $p(x)$ "recursivamente" en función de la función de masa de probabilidad evaluada con los valores más pequeños $x - 3, x - 4, \ldots, 2$.] (Este problema fue sugerido del artículo "Evaluation of a Start-Up Demonstration Test", *J. Quality Technology*, 1983: 103–106.)

106. Una aerolínea ha desarrollado un plan para un club de viajeros ejecutivos sobre la premisa de que 10% de sus clientes actuales calificarían para la membresía.

a. Suponiendo la validez de esta premisa, entre 25 clientes actuales seleccionados al azar, ¿cuál es la probabilidad de que entre 2 y 6 (inclusive) califiquen para la membresía?

b. De nuevo suponiendo la validez de la premisa, ¿cuál es el número esperado de clientes que califican y la desviación estándar del número que califica en una muestra aleatoria de 100 clientes actuales?

c. Sea X el número en una muestra al azar de 25 clientes actuales que califican para la membresía. Considere rechazar la premisa de la compañía a favor de la pretensión de que $p > .10$ si $x \geq 7$. ¿Cuál es la probabilidad de que la premisa de la compañía sea rechazada cuando en realidad es válida?

d. Remítase a la regla de decisión introducida en el inciso (c). ¿Cuál es la probabilidad de que la premisa de la compañía no sea rechazada aun cuando $p = .20$ (es decir, 20% califican)?

107. Cuarenta por ciento de las semillas de mazorcas de maíz (maíz moderno) portan sólo una espiga y el 60% restante portan dos espigas. Una semilla con una espiga producirán una mazorca con espigas únicas 29% del tiempo, en tanto que una semilla

con dos espigas producirán una mazorca con espigas únicas 26% del tiempo. Considere seleccionar al azar diez semillas.

a. ¿Cuál es la probabilidad de que exactamente cinco de estas semillas porten una sola espiga y de que produzcan una mazorca con una sola espiga?

b. ¿Cuál es la probabilidad de que exactamente cinco de estas mazorcas producidas por estas semillas tengan espigas únicas? ¿Cuál es la probabilidad de que cuando mucho cinco mazorcas tengan espigas únicas?

108. Un juicio terminó con el jurado en desacuerdo porque ocho de sus miembros estuvieron a favor de un veredicto de culpabilidad y los otros cuatro estuvieron a favor de la absolución. Si los jurados salen de la sala en orden aleatorio y cada uno de los primeros cuatro que salen de la sala es acosado por un reportero para entrevistarlo, ¿cuál es la función de masa de probabilidad de X = el número de jurados a favor de la absolución entre los entrevistados? ¿Cuántos de los que están a favor de la absolución espera que sean entrevistados?

109. Un servicio de reservaciones emplea cinco operadores de información que reciben solicitudes de información independientemente uno de otro, cada uno de acuerdo con un proceso de Poisson con rapidez $\alpha = 2$ por minuto.

a. ¿Cuál es la probabilidad de que durante un periodo de 1 minuto dado, el primer operador no reciba solicitudes?

b. ¿Cuál es la probabilidad de que durante un periodo de 1 minuto dado, exactamente cuatro de los cinco operadores no reciban solicitudes?

c. Escriba una expresión para la probabilidad de que durante un periodo de 1 minuto dado, todos los operadores reciban exactamente el mismo número de solicitudes.

110. En un gran campo se distribuyen al azar las langostas de acuerdo con una distribución de Poisson con parámetro $\alpha = 2$ por yarda cuadrada. ¿Qué tan grande deberá ser el radio R de una región de muestreo circular para que la probabilidad de hallar por lo menos una en la región sea igual a .99?

111. Un puesto de periódicos ha pedido cinco ejemplares de cierto número de una revista de fotografía. Sea X = el número de individuos que vienen a comprar esta revista. Si X tiene una distribución de Poisson con parámetro $\mu = 4$, ¿cuál es el número esperado de ejemplares que serán vendidos?

112. Los individuos A y B comienzan a jugar una secuencia de partidas de ajedrez. Sea S = {A gana un juego} y suponga que los resultados de juegos sucesivos son independientes con $P(S) = p$ y $P(F) = 1 - p$ (nunca empatan). Jugarán hasta que uno de ellos gane diez juegos. Sea X = el número de partidas jugadas (con posibles valores 10, 11, . . . , 19).

a. Con $x = 10, 11, \ldots, 19$, obtenga una expresión para $p(x) = P(X = x)$.

b. Si un empate es posible, con $p = P(S)$, $q = P(F)$, $1 - p - q = P(\text{empate})$, ¿cuáles son los posibles valores de X? ¿Cuál es $P(20 \le X)$? [*Sugerencia*: $P(20 \le X) = 1 - P(X < 20)$.]

113. Un análisis para detectar la presencia de una enfermedad tiene una probabilidad de .20 de dar un resultado falso positivo (que indica que un individuo tiene la enfermedad cuando éste no es el caso) y una probabilidad de .10 de dar un resultado falso negativo. Suponga que diez individuos son analizados, cinco de los cuales tienen la enfermedad y cinco de los cuales no. Sea X = el número de lecturas positivas que resultan.

a. ¿Tiene X una distribución binomial? Explique su razonamiento.

b. ¿Cuál es la probabilidad de que exactamente tres de diez resultados sean positivos?

114. La función de masa de probabilidad binomial negativa generalizada está dada por

$$nb(x; r, p) = k(r, x) \cdot p^r(1 - p)^x \qquad x = 0, 1, 2, \ldots$$

Sea X el número de plantas de cierta especie encontradas en una región particular y tenga esta distribución con $p = .3$ y $r = 2.5$. ¿Cuál es $P(X = 4)$? ¿Cuál es la probabilidad de que por lo menos se encuentre una planta?

115. Hay dos contadores públicos en una oficina particular que preparan declaraciones de impuestos para los clientes. Supongamos que para un tipo particular de forma compleja, el número de errores cometidos por el preparador de la primera tiene una distribución de Poisson con media μ_1, el número de errores cometidos por el preparador de la segunda tiene una distribución de Poisson con media μ_2 y que cada contador prepara el mismo número de formas de este tipo. Entonces, si una forma de este tipo es seleccionada al azar, la función

$$p(x; \mu_1, \mu_2) = .5\frac{e^{-\mu_1}\mu_1^x}{x!} + .5\frac{e^{-\mu_2}\mu_2^x}{x!} \qquad x = 0, 1, 2, \ldots$$

da la función de masa de probabilidad de X = el número de errores en el formulario seleccionado.

a. Compruebe que $p(x; \mu_1, \mu_2)$ es de hecho una función de masa de probabilidad legítima (≥ 0 y se suma a 1).

b. ¿Cuál es el número esperado de errores en el formulario seleccionado?

c. ¿Cuál es la varianza del número de errores en el formulario seleccionado?

d. ¿Cómo cambia la función de masa de probabilidad si el primer contador prepara el 60% de todas esas formas y el segundo prepara el 40%?

116. La *moda* de una variable aleatoria discreta X con función de masa de probabilidad $p(x)$ es ese valor x^* con el cual $p(x)$ alcanza su valor más grande (el valor x más probable).

a. Sea $X \sim \text{Bin}(n, p)$. Considerando la razón $b(x + 1; n, p)/b(x; n, p)$, demuestre que $b(x; n, p)$ se incrementa con x en tanto $x < np - (1 - p)$. Concluya que la moda x^* es el entero que satisface $(n + 1)p - 1 \le x^* \le (n + 1)p$.

b. Demuestre que si X tiene una distribución de Poisson con parámetro μ, la moda es el entero más grande menor que μ. Si μ es un entero, demuestre que tanto $\mu - 1$ como μ son modas.

117. Un disco duro de computadora tiene diez pistas concéntricas, numeradas 1, 2, . . . , 10 desde la más externa hasta la más interna y un solo brazo de acceso. Sea p_i = la probabilidad de que cualquier solicitud particular de datos hará que el brazo se vaya a la pista i ($i = 1, \ldots, 10$). Suponga que las pistas recorridas en búsquedas sucesivas son independientes. Sea X = el número de pistas sobre las cuales pasa el brazo de acceso durante dos solicitudes sucesivas (excluida la pista que el brazo acaba de dejar, así que los valores posibles son $x = 0$,

1, . . . , 9). Calcule la función de masa de probabilidad de X. [*Sugerencia*: P(el brazo está ahora sobre la pista i y $X = j$) = $P(X = j|$el brazo está ahora sobre $i) \cdot p_i$. Una vez que se escribe la probabilidad condicional en función de $p_1, \ldots, p_{10}$, mediante la ley de probabilidad total, se obtiene la probabilidad deseada sumando a lo largo de i.]

118. Si X es una variable aleatoria hipergeométrica demuestre directamente con la definición que $E(X) = nM/N$ (considere sólo el caso $n < M$). [*Sugerencia*: saque como factor a nM/N de la suma para $E(X)$ y demuestre que los términos adentro de la suma son de la forma $h(y; n - 1, M - 1, N - 1)$ donde $y = x - 1$.]

119. Use el hecho de que

$$\sum_{\text{toda } x} (x - \mu)^2 p(x) \geq \sum_{x:\, |x-\mu| \geq k\sigma} (x - \mu)^2 p(x)$$

para comprobar la desigualdad de Chebyshev dada en el ejercicio 44.

120. El proceso de Poisson simple de la sección 3.6 está caracterizado por una rapidez constante α a la cual los eventos ocurren por unidad de tiempo. Una generalización de esto es suponer que la probabilidad de que ocurra exactamente un evento en el intervalo $[t, t + \Delta t]$ es $\alpha(t) \cdot \Delta t + o(\Delta t)$. Se puede demostrar entonces que el número de eventos que ocurren durante un intervalo $[t_1, t_2]$ tiene una distribución de Poisson con parámetro

$$\mu = \int_{t_2}^{t_1} \alpha(t)\, dt$$

La ocurrencia de eventos en el transcurso del tiempo en esta situación se llama *proceso de Poisson no homogéneo*. El artículo "Inference Based on Retrospective Ascertainment", *J. Amer. Stat. Assoc.*, 1989: 360–372, considera la función de intensidad

$$\alpha(t) = e^{a+bt}$$

apropiada para eventos que implican la transmisión de VIH (el virus del SIDA) vía transfusiones sanguíneas. Suponga que $a = 2$ y $b = .6$ (cercanos a los valores sugeridos en el artículo), con el tiempo en años.

a. ¿Cuál es el número esperado de eventos en el intervalo [0, 4]? ¿En [2, 6]?

b. ¿Cuál es la probabilidad de que cuando mucho ocurran 15 eventos en el intervalo [0, .9907]?

121. Considere un conjunto de $A_1, \ldots, A_k$ de eventos mutuamente exclusivos y exhaustivos y una variable aleatoria X cuya distribución depende de cuál de los eventos A_i ocurra (p. ej., un viajero podría seleccionar una de tres rutas posibles de su casa al trabajo, con X como el tiempo de recorrido). Sea $E(X|A_i)$ el valor esperado de X dado que el evento A_i ocurre. Entonces se puede demostrar que $E(X) = \Sigma E(X|A_i) \cdot P(A_i)$ es el promedio ponderado de las "expectativas condicionales" individuales donde las ponderaciones son las probabilidades de la división de eventos.

a. La duración esperada de una llamada de voz a un número telefónico particular es de 3 minutos, mientras que la duración esperada de una llamada de datos a ese mismo número es de 1 minuto. Si 75% de las llamadas son de voz, ¿cuál es la duración esperada de la siguiente llamada?

b. Una pastelería vende tres diferentes tipos de galletas con chispas de chocolate. El número de chispas de chocolate en un tipo i de galleta tiene una distribución de Poisson con parámetro $\mu_i = i + 1 \quad (i = 1, 2, 3)$. Si 20% de todos los clientes que compran una galleta con chispas de chocolate selecciona el primer tipo, 50% elige el segundo tipo y el 30% restante opta por el tercer tipo, ¿cuál es el número esperado de chispas en una galleta comprada por el siguiente cliente?

122. Considere una fuente de comunicaciones que transmite paquetes que contienen lenguaje digitalizado. Después de cada transmisión, el receptor envía un mensaje que indica si la transmisión fue exitosa o no. Si una transmisión no es exitosa, el paquete es reenviado. Suponga que el paquete de voz puede ser transmitido un máximo de 10 veces. Suponiendo que los resultados de transmisiones sucesivas son independientes uno de otro y que la probabilidad de que cualquier transmisión particular sea exitosa es p, determine la función de masa de probabilidad de la variable aleatoria X = el número de veces que un paquete es transmitido. Luego obtenga una expresión para el número de veces que se espera que un paquete sea transmitido.

Bibliografía

Johnson, Norman, Samuel Kotz y Adrienne Kemp, *Discrete Univariate Distributions*. Wiley, Nueva York, 1992. Una enciclopedia de información sobre distribuciones discretas.

Olkin, Ingram, Cyrus Derman y Leon Gleser, *Probability Models and Applications* (2a. ed.), Macmillan, Nueva York, 1994. Contiene una discusión a fondo tanto de las propiedades generales de distribuciones discretas y continuas como los resultados para distribuciones específicas.

Ross, Sheldon, *Introduction to Probability Models* (9a. ed.), Academic Press, Nueva York, 2007. Una buena fuente de material sobre el proceso de Poisson y generalizaciones, y una amena introducción a otros temas de probabilidad aplicada.

4 Variables aleatorias continuas y distribuciones de probabilidad

INTRODUCCIÓN

El capítulo 3 se concentró en el desarrollo de distribuciones de probabilidad de variables aleatorias discretas. En este capítulo se estudia el segundo tipo general de variable aleatoria que se presenta en muchos problemas aplicados. Las secciones 4.1 y 4.2 presentan las definiciones y propiedades básicas de las variables aleatorias continuas y sus distribuciones de probabilidad. En la sección 4.3 se estudia con detalle la variable aleatoria normal y su distribución, sin duda la más importante y útil en la probabilidad y estadística. Las secciones 4.4 y 4.5 se ocupan de otras distribuciones continuas utilizadas con frecuencia en trabajo aplicado. En la sección 4.6 se introduce un método de evaluar si un dato muestral es compatible con una distribución especificada.

4.1 Funciones de densidad de probabilidad

Una variable aleatoria discreta es una cuyos valores posibles constituyen un conjunto finito o bien pueden ser puestos en lista en una secuencia infinita (una lista en la cual existe un primer elemento, un segundo elemento, etc.). Una variable aleatoria cuyo conjunto de valores posibles es un intervalo completo de números no es discreta.

De acuerdo con el capítulo 3 recuérdese que una variable aleatoria X es continua si (1) sus valores posibles comprenden un solo intervalo sobre la línea de numeración (para alguna $A < B$, cualquier número x entre A y B es un valor posible) o una unión de intervalos disjuntos y (2) $P(X = c) = 0$ para cualquier número c que sea un valor posible de X.

Ejemplo 4.1 En el estudio de la ecología de un lago, se mide la profundidad en lugares seleccionados, entonces $X =$ la profundidad en ese lugar es una variable aleatoria continua. En este caso A es la profundidad mínima en la región muestreada y B es la profundidad máxima. ■

Ejemplo 4.2 Si se selecciona al azar un compuesto químico y se determina su pH X, entonces X es una variable aleatoria continua porque cualquier valor pH entre 0 y 14 es posible. Si se conoce más sobre el compuesto seleccionado para su análisis, entonces el conjunto de posibles valores podría ser un subintervalo de [0, 14], tal como $5.5 \leq x \leq 6.5$, pero X seguiría siendo continua. ■

Ejemplo 4.3 Sea X la cantidad de tiempo que un cliente seleccionado al azar pasa esperando antes de que comience su corte de pelo. El primer pensamiento podría ser que X es una variable aleatoria continua, puesto que se requiere medirla para determinar su valor. Sin embargo, existen clientes suficientemente afortunados que no tienen que esperar antes de sentarse en el sillón del peluquero. Así que el caso debe ser $P(X = 0) > 0$. Aunque, en caso de que no haya sillones vacíos, el tiempo de espera será continuo puesto que X podría asumir entonces cualquier valor entre un tiempo mínimo posible A y un tiempo máximo posible B. Esta variable aleatoria no es ni puramente discreta ni puramente continua sino que es una mezcla de los dos tipos. ■

Se podría argumentar que aunque en principio variables tales como altura, peso y temperatura son continuas, en la práctica las limitaciones de los instrumentos de medición nos restringen a un mundo discreto (aunque en ocasiones muy finamente subdividido). Sin embargo, los modelos continuos a menudo representan muy bien de forma aproximada situaciones del mundo real y con frecuencia es más fácil trabajar con matemáticas continuas (el cálculo) que con matemáticas de variables discretas y distribuciones.

Distribuciones de probabilidad de variables continuas

Supóngase que la variable X de interés es la profundidad de un lago en un punto sobre la superficie seleccionado al azar. Sea $M =$ la profundidad máxima (en metros), así que cualquier número en el intervalo $[0, M]$ es un valor posible de X. Si se "discretiza" X midiendo la profundidad al metro más cercano, entonces los valores posibles son enteros no negativos menores o iguales a M. La distribución discreta de profundidad resultante se ilustra con un histograma de probabilidad. Si se traza el histograma de modo que el área del rectángulo sobre cualquier entero posible k sea la proporción del lago cuya profundidad es (al metro más cercano) k, entonces el área total de todos los rectángulos es 1. En la figura 4.1(a) aparece un posible histograma.

Si se mide la profundidad con mucho más precisión y se utiliza el mismo eje de medición de la figura 4.1(a), cada rectángulo en el histograma de probabilidad resultante es mucho más angosto, aun cuando el área total de todos los rectángulos sigue siendo 1.

En la figura 4.1(b) se ilustra un posible histograma; tiene una apariencia mucho más regular que el histograma de la figura 4.1(a). Si se continúa de esta manera midiendo la profundidad más y más finamente, la secuencia resultante de histogramas se aproxima a una curva más regular, tal como la ilustrada en la figura 4.1(c). Como en cada histograma el área total de todos los rectángulos es igual a 1, el área total bajo la curva regular también es 1. La probabilidad de que la profundidad en un punto seleccionado al azar se encuentre entre a y b es simplemente el área bajo la curva regular entre a y b. Es de manera exacta una curva suave del tipo ilustrado en la figura 4.1(c) la que especifica una distribución de probabilidad continua.

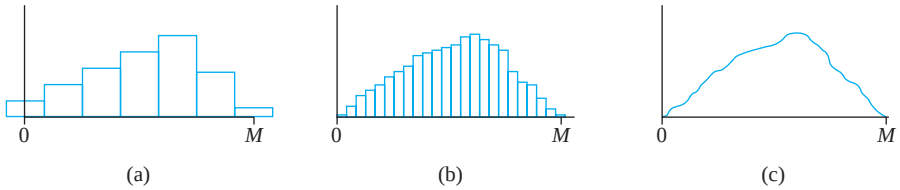


Figura 4.1 (a) Histograma de probabilidad de profundidad medida al metro más cercano; (b) histograma de probabilidad de profundidad medida al centímetro más cercano; (c) un límite de una secuencia de histogramas discretos

DEFINICIÓN

Sea X una variable aleatoria continua. Entonces, una **distribución de probabilidad** o **función de densidad de probabilidad** (fdp) de X es una función $f(x)$ tal que para dos números cualesquiera a y b con $a \le b$,

$$P(a \le X \le b) = \int_a^b f(x)dx$$

Es decir, la probabilidad de que X asuma un valor en el intervalo $[a, b]$ es el área sobre este intervalo y bajo la gráfica de la función de densidad, como se ilustra en la figura 4.2. La gráfica de $f(x)$ a menudo se conoce como *curva de densidad*.

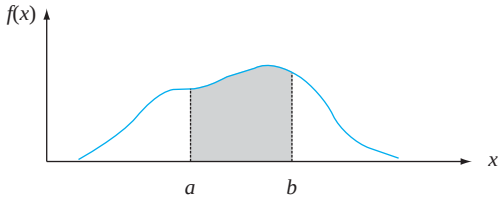


Figura 4.2 $P(a \le X \le b) =$ el área bajo la curva de densidad entre a y b

Para que $f(x)$ sea una función de densidad de probabilidad legítima, debe satisfacer las dos siguientes condiciones:

1. $f(x) \ge 0$ con todas las x

2. $\int_{-\infty}^{\infty} f(x)\,dx =$ área bajo toda la gráfica de $f(x)$
$= 1$

Ejemplo 4.4 La dirección de una imperfección con respecto a una línea de referencia sobre un objeto circular como un neumático, un rotor de freno o un volante está, en general, sujeta a incertidumbre. Considérese la línea de referencia que conecta el vástago de la válvula de un neumático con el punto central y sea X el ángulo medido en el sentido de las manecillas

del reloj con respecto a la ubicación de una imperfección. Una posible función de densidad de probabilidad de X es

$$f(x) = \begin{cases} \frac{1}{360} & 0 \le x < 360 \\ 0 & \text{de lo contrario} \end{cases}$$

La función de densidad de probabilidad aparece graficada en la figura 4.3. Claramente $f(x) \ge 0$. El área bajo la curva de densidad es simplemente el área de un rectángulo: (altura)(base) $= \left(\frac{1}{360}\right)(360) = 1$. La probabilidad de que el ángulo sea de entre 90° y 180° es

$$P(90 \le X \le 180) = \int_{90}^{180} \frac{1}{360}\, dx = \frac{x}{360}\Big|_{x=90}^{x=180} = \frac{1}{4} = .25$$

La probabilidad de que el ángulo de ocurrencia esté dentro de 90° de la línea de referencia es

$$P(0 \le X \le 90) + P(270 \le X < 360) = .25 + .25 = .50$$

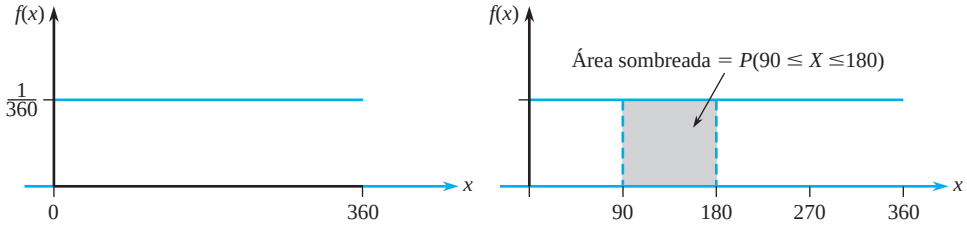


Figura 4.3 Función de densidad de probabilidad del ejemplo 4.4 ■

Como siempre que $0 \le a \le b \le 360$ en el ejemplo 4.4 y $P(a \le X \le b)$ depende sólo del ancho $b - a$ del intervalo, se dice que X tiene una distribución uniforme.

DEFINICIÓN

Se dice que una variable aleatoria continua X tiene una **distribución uniforme** en el intervalo $[A, B]$ si la función de densidad de probabilidad de X es

$$f(x; A, B) = \begin{cases} \frac{1}{B - A} & A \le x \le B \\ 0 & \text{de lo contrario} \end{cases}$$

La gráfica de cualquier función de densidad de probabilidad uniforme es como la de la figura 4.3, excepto que el intervalo de densidad positiva es $[A, B]$ en lugar de $[0, 360]$.

En el caso discreto, una función de masa de probabilidad (fmp) dice cuántas pequeñas "burbujas" de masa de probabilidad de varias magnitudes están distribuidas a lo largo del eje de medición. En el caso continuo, la densidad de probabilidad está "repartida" en forma continua a lo largo del intervalo de posibles valores. Cuando la densidad está distribuida uniformemente a lo largo del intervalo, se obtiene una función de densidad de probabilidad uniforme como en la figura 4.3.

Cuando X es una variable aleatoria discreta, a cada valor posible se le asigna una probabilidad positiva. Esto no es cierto en el caso de una variable aleatoria continua (es decir,

se satisface la segunda condición de la definición) porque el área bajo una curva de densidad situada sobre cualquier valor único es cero:

$$P(X = c) = \int_{c}^{c} f(x)\,dx = \lim_{\varepsilon \to 0} \int_{c-\varepsilon}^{c+\varepsilon} f(x)\,dx = 0$$

El hecho de que $P(X = c) = 0$ cuando X es continua tiene una importante consecuencia práctica: la probabilidad de que X quede en algún intervalo entre a y b no depende de si el límite inferior a o el límite superior b está incluido en el cálculo de probabilidad:

$$P(a \le X \le b) = P(a < X < b) = P(a < X \le b) = P(a \le X < b) \qquad (4.1)$$

Si X es discreta y tanto a como b son valores posibles (p. ej., X es binomial con $n = 20$ y $a = 5$, $b = 10$), entonces todas las cuatro probabilidades en (4.1) son diferentes.

La condición de probabilidad cero tiene un análogo físico. Considérese una barra circular sólida con área de sección transversal = 1 pulg^2. Coloque la barra a lo largo de un eje de medición y supóngase que la densidad de la barra en cualquier punto x está dada por el valor $f(x)$ de una función de densidad. Entonces si la barra se rebana en los puntos a y b y este segmento se retira, la cantidad de masa eliminada es $\int_a^b f(x)\,dx$; si la barra se rebana exactamente en el punto c, no se elimina masa. Se asigna masa a segmentos de intervalo de la barra pero no a puntos individuales.

Ejemplo 4.5 "Intervalo de tiempo" en el flujo de tránsito es el tiempo transcurrido entre el tiempo en que un carro termina de pasar por un punto fijo y el instante en que el siguiente carro comienza a pasar por ese punto. Sea X = el intervalo de tiempo para dos carros consecutivos seleccionados al azar en una autopista durante un periodo de tráfico intenso. La siguiente función de densidad de probabilidad de X es en esencia el sugerido en "The Statistical Properties of Freeway Traffic" (*Transp. Res.* vol. 11: 221−228):

$$f(x) = \begin{cases} .15e^{-.15(x-.5)} & x \ge .5 \\ 0 & \text{de lo contrario} \end{cases}$$

La gráfica de $f(x)$ se da en la figura 4.4; no hay ninguna densidad asociada con intervalos de tiempo de menos de .5 y la densidad del intervalo de tiempo decrece con rapidez (exponencialmente rápido) a medida que x se incrementa a partir de .5. Claramente, $f(x) \ge 0$; para demostrar que $\int_{-\infty}^{\infty} f(x)\,dx = 1$, se utiliza el resultado obtenido con cálculo integral $\int_a^{\infty} e^{-kx}\,dx = (1/k)e^{-k \cdot a}$. Entonces

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{.5}^{\infty} .15e^{-.15(x-.5)}\,dx = .15e^{.075}\int_{.5}^{\infty} e^{-.15x}\,dx$$
$$= .15e^{.075} \cdot \frac{1}{.15}\,e^{-(.15)(.5)} = 1$$

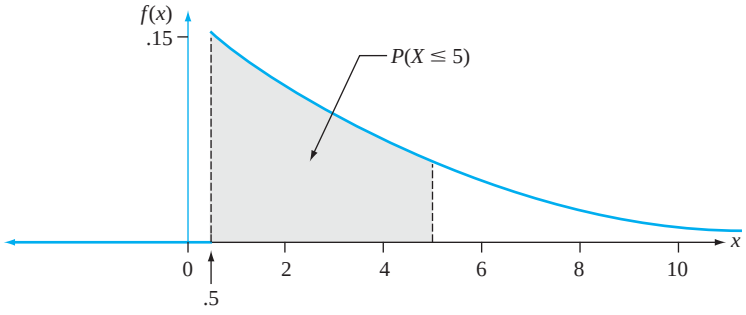


Figura 4.4 Curva de densidad del intervalo de tiempo entre vehículos en el ejemplo 4.5

La probabilidad de que el intervalo de tiempo sea cuando mucho de 5 segundos es

$$\begin{aligned} P(X \le 5) &= \int_{-\infty}^{5} f(x)\,dx = \int_{.5}^{5} .15e^{-.15(x-.5)}\,dx \\ &= .15e^{.075}\int_{.5}^{5} e^{-.15x}\,dx = .15e^{.075}\cdot\left(-\frac{1}{.15}e^{-.15x}\Big|_{x=.5}^{x=5}\right) \\ &= e^{.075}(-e^{-.75} + e^{-.075}) = 1.078(-.472 + .928) = .491 \\ &= P(\text{menos de 5 s}) = P(X < 5) \end{aligned}$$

■

A diferencia de las distribuciones discretas como la binomial, la hipergeométrica y la binomial negativa, la distribución de cualquier variable aleatoria continua dada no puede, en general, ser obtenida mediante argumentos probabilísticos. En cambio, se debe hacer una selección juiciosa de la función de densidad de probabilidad basada en conocimientos previos y en los datos disponibles. Afortunadamente, existen algunas familias generales de funciones de densidad de probabilidad que se ajustan bien a una amplia variedad de situaciones experimentales; varias de éstas se discuten más adelante en el capítulo.

Exactamente como en el caso discreto, a menudo es útil pensar en la población de interés como compuesta de valores X en lugar de individuos u objetos. La función de densidad de probabilidad es entonces un modelo de la distribución de valores en esta población numérica y con base en este modelo se pueden calcular varias características de la población (tal como la media).

EJERCICIOS Sección 4.1 (1−10)

1. La corriente en un circuito determinado, medido por un amperímetro es una variable aleatoria continua X con la función de densidad siguiente:

$$f(x) = \begin{cases} .075x + .2 & 3 \le x \le 5 \\ 0 & \text{de lo contrario} \end{cases}$$

a. Grafique la función de densidad de probabilidad para verificar que el área total bajo la curva de densidad es de hecho 1.
b. Calcule $P(X \le 4)$. ¿Cómo se compara esta probabilidad con $P(X < 4)$?
c. Calcule $P(3.5 \le X \le 4.5)$ y $P(4.5 < X)$

2. Suponga que la temperatura de reacción X (en °C) en cierto proceso químico tiene una distribución uniforme con $A = -5$ y $B = 5$.
a. Calcule $P(X < 0)$.
b. Calcule $P(-2.5 < X < 2.5)$.
c. Calcule $P(-2 \le X \le 3)$.
d. Para que k satisfaga $-5 < k < k + 4 < 5$, calcule $P(k < X < k + 4)$.

3. El error implicado al hacer una medición es una variable aleatoria continua X con función de densidad de probabilidad

$$f(x) = \begin{cases} .09375(4 - x^2) & -2 \le x \le 2 \\ 0 & \text{de lo contrario} \end{cases}$$

a. Trace la gráfica de $f(x)$.
b. Calcule $P(X > 0)$.
c. Calcule $P(-1 < X < 1)$.
d. Calcule $P(X < -.5 \text{ o } X > .5)$.

4. Sea X el esfuerzo vibratorio (lb/pulg^2) en el aspa de una turbina de viento a una velocidad del viento particular en un túnel aerodinámico. El artículo "Blade Fatigue Life Assessment with Application to VAWTS" (*J. of Solar Energy Engr.*, 1982: 107–111) propone la distribución de Rayleigh, con función de densidad de probabilidad

$$f(x;\theta) = \begin{cases} \dfrac{x}{\theta^2}\cdot e^{-x^2/(2\theta^2)} & x > 0 \\ 0 & \text{de lo contrario} \end{cases}$$

como modelo de la distribución X.
a. Verifique que $f(x; \theta)$ es una función de densidad de probabilidad legítima.
b. Suponga que $\theta = 100$ (un valor sugerido por una gráfica en el artículo). ¿Cuál es la probabilidad de que X sea cuando mucho de 200? ¿Menos de 200? ¿Por lo menos de 200?
c. ¿Cuál es la probabilidad de que X esté entre 100 y 200 (de nuevo suponiendo $\theta = 100$)?
d. Dé una expresión para $P(X \le x)$.

5. Un profesor universitario nunca termina su disertación antes del final de la hora y siempre termina dentro de dos minutos después de la hora. Sea X = el tiempo que transcurre entre el final de la hora y el final de la disertación y suponga que la función de densidad de probabilidad de X es

$$f(x) = \begin{cases} kx^2 & 0 \le x \le 2 \\ 0 & \text{de lo contrario} \end{cases}$$

a. Determine el valor de k y trace la curva de densidad correspondiente. [*Sugerencia*: el área total bajo la gráfica de $f(x)$ es 1.]
b. ¿Cuál es la probabilidad de que la disertación termine dentro de 1 minuto del final de la hora?

c. ¿Cuál es la probabilidad de que la disertación continúe después de la hora durante entre 60 y 90 segundos?
d. ¿Cuál es la probabilidad de que la disertación continúe durante por lo menos 90 segundos después del final de la hora?

6. El peso real de lectura de una pastilla de estéreo ajustado a 3 gramos en un tocadiscos particular puede ser considerado como una variable aleatoria continua X con función de densidad de probabilidad

$$f(x) = \begin{cases} k[1 - (x - 3)^2] & 2 \le x \le 4 \\ 0 & \text{de lo contrario} \end{cases}$$

a. Trace la gráfica de $f(x)$.
b. Determine el valor de k.
c. ¿Cuál es la probabilidad de que el peso real de lectura sea mayor que el peso prescrito?
d. ¿Cuál es la probabilidad de que el peso real de lectura esté dentro de .25 gramo del peso prescrito?
e. ¿Cuál es la probabilidad de que el peso real difiera del peso prescrito por más de .5 gramo?

7. Se cree que el tiempo X (minutos) para que un ayudante de laboratorio prepare el equipo para cierto experimento tiene una distribución uniforme con $A = 25$ y $B = 35$.
a. Determine la función de densidad de probabilidad de X y trace la curva de densidad correspondiente.
b. ¿Cuál es la probabilidad de que el tiempo de preparación exceda de 33 minutos?
c. ¿Cuál es la probabilidad de que el tiempo de preparación esté dentro de 2 min del tiempo medio? [*Sugerencia*: identifique μ en la gráfica de $f(x)$.]
d. Para cualquier a tal que $25 < a < a + 2 < 35$, ¿cuál es la probabilidad de que el tiempo de preparación sea de entre a y $a + 2$ minutos?

8. Para ir al trabajo, un profesor primero debe tomar un camión cerca de su casa y luego tomar un segundo camión. Si el tiempo de espera (en minutos) en cada parada tiene una distribución uniforme con $A = 0$ y $B = 5$, entonces se puede demostrar que el tiempo de espera total Y tiene la función de densidad de probabilidad

$$f(y) = \begin{cases} \frac{1}{25}y & 0 \le y < 5 \\ \frac{2}{5} - \frac{1}{25}y & 5 \le y \le 10 \\ 0 & y < 0 \text{ o } y > 10 \end{cases}$$

a. Trace la gráfica de la función de densidad de probabilidad de Y.
b. Verifique que $\int_{-\infty}^{\infty} f(y)\,dy = 1$.
c. ¿Cuál es la probabilidad de que el tiempo de espera total sea cuando mucho de 3 min?
d. ¿Cuál es la probabilidad de que el tiempo de espera total sea cuando mucho de 8 min?
e. ¿Cuál es la probabilidad de que el tiempo de espera total sea de entre 3 y 8 min?
f. ¿Cuál es la probabilidad de que el tiempo de espera total sea de menos de 2 min o de más de 6 min?

9. Considere de nuevo la función de densidad de probabilidad de X = intervalo de tiempo dado en el ejemplo 4.5. ¿Cuál es la probabilidad de que el intervalo de tiempo sea
a. cuando mucho de 6 segundos?
b. de más de 6 segundos? ¿Por lo menos de 6 segundos?
c. de entre 5 y 6 segundos?

10. Una familia de funciones de densidad de probabilidad que ha sido utilizada para aproximar la distribución del ingreso, el tamaño de la población de una ciudad y el tamaño de compañías es la familia Pareto. La familia tiene dos parámetros, k y θ, ambos > 0, y la función de densidad de probabilidad es

$$f(x; k, \theta) = \begin{cases} \frac{k \cdot \theta^k}{x^{k+1}} & x \ge \theta \\ 0 & x < \theta \end{cases}$$

a. Trace la gráfica de $f(x; k, \theta)$.
b. Verifique que el área total bajo la gráfica es igual a 1.
c. Si la variable aleatoria X tiene una función de densidad de probabilidad $f(x; k, \theta)$, con cualquier $b > \theta$ fija, obtenga una expresión para $P(X \le b)$.
d. Para $\theta < a < b$, obtenga una expresión para la probabilidad $P(a \le X \le b)$.

4.2 Funciones de distribución acumulativa y valores esperados

Varios de los más importantes conceptos introducidos en el estudio de distribuciones discretas también desempeñan un importante papel en las distribuciones continuas. Definiciones análogas a las del capítulo 3 implican reemplazar la suma por integración.

Función de distribución acumulativa

La función de distribución acumulativa $F(x)$ de una variable aleatoria discreta X da, con cualquier número especificado x, la probabilidad $P(X \le x)$. Se obtiene sumando la función de masa de probabilidad $p(y)$ a lo largo de todos los valores posibles y que satisfagan $y \le x$. La función de distribución acumulativa de una variable aleatoria continua da las mismas probabilidades $P(X \le x)$ y se obtiene integrando la función de densidad de probabilidad $f(y)$ entre los límites $-\infty$ y x.

DEFINICIÓN

La **función de distribución acumulativa** $F(x)$ de una variable aleatoria continua X se define para todo número x como

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(y)\,dy$$

Para cada x, $F(x)$ es el área bajo la curva de densidad a la izquierda de x. Esto se ilustra en la figura 4.5, donde $F(x)$ se incrementa con suavidad a medida que x se incrementa.

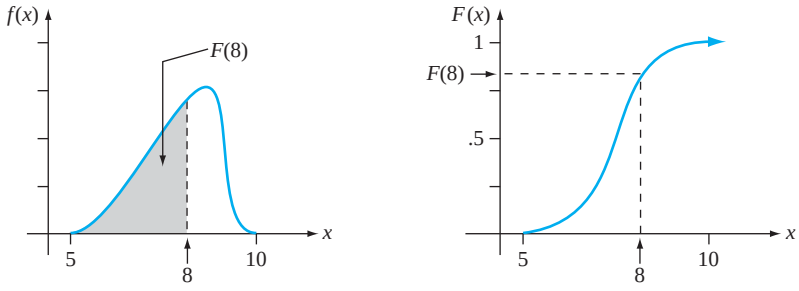


Figura 4.5 Una función de densidad de probabilidad y función de distribución acumulativa asociada

Ejemplo 4.6 Sea X el espesor de una cierta lámina de metal con distribución uniforme en $[A, B]$. La función de densidad se muestra en la figura 4.6. Para $x < A$, $F(x) = 0$, dado que no hay área bajo la gráfica de la función de densidad a la izquierda de la x. Con $x \ge B$, $F(x) = 1$, puesto que toda el área está acumulada a la izquierda de la x. Finalmente para $A \le x \le B$,

$$F(x) = \int_{-\infty}^{x} f(y)dy = \int_{A}^{x} \frac{1}{B-A}\,dy = \frac{1}{B-A} \cdot y \Big|_{y=A}^{y=x} = \frac{x-A}{B-A}$$

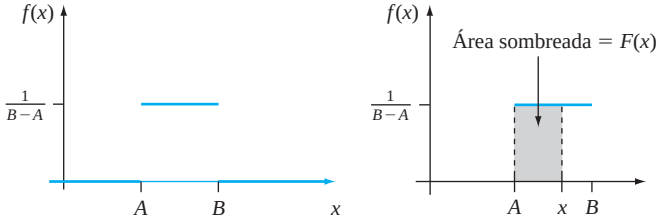


Figura 4.6 Función de densidad de probabilidad de una distribución uniforme

La función de distribución acumulativa completa es

$$F(x) = \begin{cases} 0 & x < A \\ \dfrac{x-A}{B-A} & A \le x < B \\ 1 & x \ge B \end{cases}$$

La gráfica de esta función de distribución acumulativa aparece en la figura 4.7.

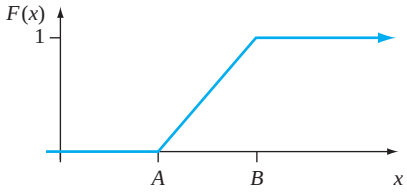


Figura 4.7 Función de distribución acumulativa de una distribución uniforme ■

Utilización de $F(x)$ para calcular probabilidades

La importancia de la función de distribución acumulativa en este caso, lo mismo que para variables aleatorias discretas, es que las probabilidades de varios intervalos pueden ser calculadas con una fórmula o tabla de $F(x)$.

PROPOSICIÓN

Sea X una variable aleatoria continua con función de densidad de probabilidad $f(x)$ y función de distribución acumulativa $F(x)$. Entonces para cualquier número a,

$$P(X > a) = 1 - F(a)$$

y para dos números cualesquiera a y b con $a < b$,

$$P(a \le X \le b) = F(b) - F(a)$$

La figura 4.8 ilustra la segunda parte de esta proposición; la probabilidad deseada es el área sombreada bajo la curva de densidad entre a y b y es igual a la diferencia entre las dos áreas acumulativas sombreadas. Esto es diferente de lo que es apropiado para una variable aleatoria discreta de valor entero (p. ej., binomial o Poisson): $P(a \le X \le b) = F(b) - F(a - 1)$ cuando a y b son enteros.

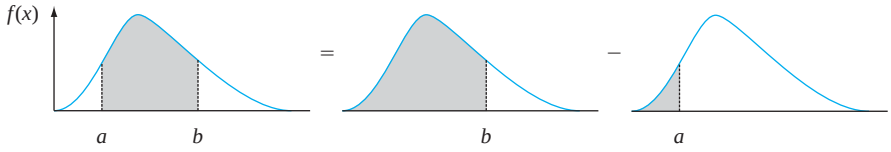


Figura 4.8 Cálculo de $P(a \le X \le b)$ a partir de probabilidades acumulativas

Ejemplo 4.7 Suponga que la función de densidad de probabilidad de la magnitud X de una carga dinámica sobre un puente (en newtons) está dada por

$$f(x) = \begin{cases} \frac{1}{8} + \frac{3}{8}x & 0 \le x \le 2 \\ 0 & \text{de lo contrario} \end{cases}$$

Para cualquier número x entre 0 y 2,

$$F(x) = \int_{-\infty}^{x} f(y)\,dy = \int_{0}^{x} \left(\frac{1}{8} + \frac{3}{8}y \right) dy = \frac{x}{8} + \frac{3}{16}x^2$$

Por lo tanto

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{8} + \frac{3}{16}x^2 & 0 \le x \le 2 \\ 1 & 2 < x \end{cases}$$

Las gráficas de $f(x)$ y $F(x)$ se muestran en la figura 4.9. La probabilidad de que la carga sea de entre 1 y 1.5 es

$$\begin{aligned} P(1 \le X \le 1.5) &= F(1.5) - F(1) \\ &= \left[\frac{1}{8}(1.5) + \frac{3}{16}(1.5)^2\right] - \left[\frac{1}{8}(1) + \frac{3}{16}(1)^2\right] \\ &= \frac{19}{64} = .297 \end{aligned}$$

La probabilidad de que la carga sea de más de 1 es

$$P(X > 1) = 1 - P(X \le 1) = 1 - F(1) = 1 - \left[\frac{1}{8}(1) + \frac{3}{16}(1)^2\right]$$

$$= \frac{11}{16} = .688$$

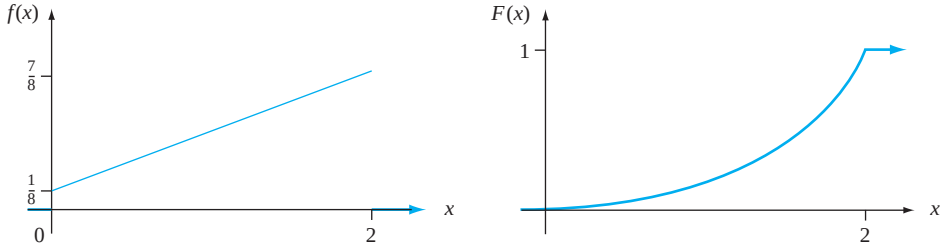


Figura 4.9 Función de densidad de probabilidad y función de distribución acumulativa del ejemplo 4.7 ■

Una vez que se obtiene la función de distribución acumulativa, cualquier probabilidad que implique X es fácil de calcular sin alguna integración adicional.

Obtención de $f(x)$ a partir de $F(x)$

Para X discreta, la función de masa de probabilidad se obtiene a partir de la función de distribución acumulativa considerando la diferencia entre dos valores $F(x)$. El análogo continuo de una diferencia es una derivada. El siguiente resultado es una consecuencia del teorema fundamental del cálculo.

PROPOSICIÓN

Si X es una variable aleatoria continua con función de densidad de probabilidad $f(x)$ y función de distribución acumulativa $F(x)$, entonces en cada x que hace posible que la derivada $F'(x)$ exista, $F'(x) = f(x)$.

Ejemplo 4.8 (Continuación del ejemplo 4.6)

Cuando X tiene una distribución uniforme, $F(x)$ es diferenciable excepto en $x = A$ y $x = B$, donde la gráfica de $F(x)$ tiene esquinas afiladas. Como $F(x) = 0$ para $x < A$ y $F(x) = 1$ para $x > B$, $F'(x) = 0 = f(x)$ con dicha x. Para $A < x < B$,

$$F'(x) = \frac{d}{dx}\left(\frac{x - A}{B - A}\right) = \frac{1}{B - A} = f(x)$$ ■

Percentiles de una distribución continua

Cuando se dice que la calificación de un individuo en una prueba estaba en el 85º percentil de la población, significa que el 85% de todas las calificaciones de la población estuvieron por debajo de dicha calificación y que el 15% estuvo arriba. Asimismo, el 40º percentil es la calificación que sobrepasa al 40% de todas las calificaciones y es superado por el 60% de todas las calificaciones.

DEFINICIÓN

Sea p un número entre 0 y 1. El **(100p)° percentil** de la distribución de una variable aleatoria continua X, denotada por $\eta(p)$, se define como

$$p = F(\eta(p)) = \int_{-\infty}^{\eta(p)} f(y)\,dy \qquad (4.2)$$

De acuerdo con la expresión (4.2), $\eta(p)$ es ese valor sobre el eje de medición de tal suerte que el 100p% del área bajo la gráfica de $f(x)$ queda a la izquierda de $\eta(p)$ y 100$(1 - p)$% queda a la derecha. Por lo tanto, $\eta(.75)$, el 75avo percentil, es tal que el área bajo la gráfica de $f(x)$ a la izquierda de $\eta(.75)$ es .75. La figura 4.10 ilustra la definición.

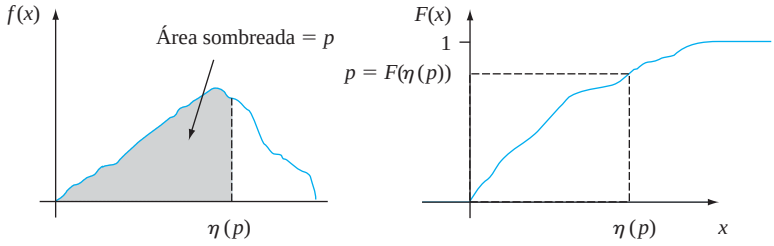


Figura 4.10 El (100p)° percentil de una distribución continua

Ejemplo 4.9 La distribución de la cantidad de grava (en toneladas) vendida por una compañía de materiales para la construcción particular en una semana dada es una variable aleatoria continua X con función de densidad de probabilidad

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & 0 \le x \le 1 \\ 0 & \text{de lo contrario} \end{cases}$$

La función de distribución acumulativa de las ventas para cualquier x entre 0 y 1 es

$$F(x) = \int_0^x \frac{3}{2}(1 - y^2)\,dy = \frac{3}{2}\left(y - \frac{y^3}{3}\right)\Big|_{y=0}^{y=x} = \frac{3}{2}\left(x - \frac{x^3}{3}\right)$$

Las gráficas tanto de $f(x)$ como de $F(x)$ aparecen en la figura 4.11. El (100p)º percentil de esta distribución satisface la ecuación

$$p = F(\eta(p)) = \frac{3}{2}\left[\eta(p) - \frac{(\eta(p))^3}{3}\right]$$

es decir,

$$(\eta(p))^3 - 3\eta(p) + 2p = 0$$

Para el 50º percentil, $p = .5$ y la ecuación que se tiene que resolver es $\eta^3 - 3\eta + 1 = 0$; la solución es $\eta = \eta(.5) = .347$. Si la distribución no cambia de una semana a otra, entonces a la larga 50% de todas las semanas se realizarán ventas de menos de .347 ton y 50% de más de .347 ton.

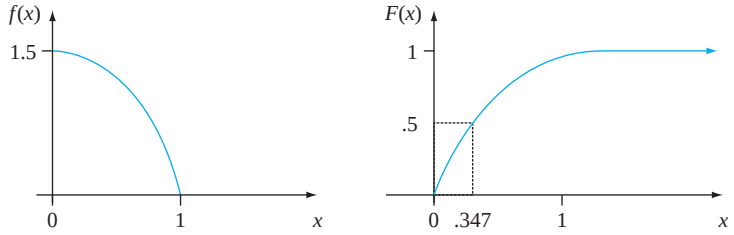


Figura 4.11 Función de densidad de probabilidad y función de distribución acumulativa del ejemplo 4.9

DEFINICIÓN

La **mediana** de una distribución continua, denotada por $\tilde{\mu}$, es el 50º percentil, así que $\tilde{\mu}$ satisface $.5 = F(\tilde{\mu})$. Es decir, la mitad del área bajo la curva de densidad se encuentra a la izquierda de $\tilde{\mu}$ y la mitad a la derecha de $\tilde{\mu}$.

Una distribución continua cuya función de densidad de probabilidad es **simétrica** —la gráfica de la función de densidad de probabilidad a la izquierda de algún punto es una imagen de espejo de la gráfica a la derecha de dicho punto—, tiene una mediana $\tilde{\mu}$ igual al punto de simetría, puesto que la mitad del área bajo la curva queda a uno u otro lado de este punto. La figura 4.12 da varios ejemplos. A menudo se supone que el error en la medición de una cantidad física tiene una distribución simétrica.

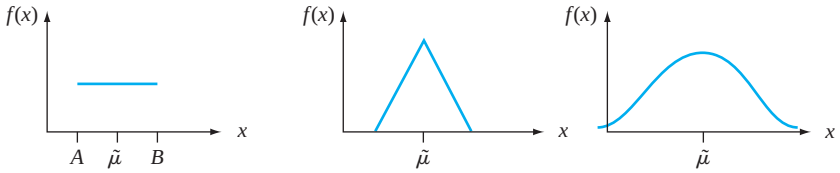


Figura 4.12 Medianas de distribuciones simétricas

Valores esperados

Para una variable aleatoria discreta X, $E(X)$ se obtuvo sumando $x \cdot p(x)$ a lo largo de posibles valores de X. Aquí se reemplaza la suma por la integración y la función de masa de probabilidad por la función de densidad de probabilidad para obtener un promedio ponderado continuo.

DEFINICIÓN

El **valor esperado** o **valor medio** de una variable aleatoria continua X con función de densidad de probabilidad $f(x)$ es

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

Ejemplo 4.10 (Continuación del ejemplo 4.9)

La función de densidad de probabilidad de las ventas semanales de grava X fue

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & 0 \le x \le 1 \\ 0 & \text{de lo contrario} \end{cases}$$

por tanto

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx = \int_{0}^{1} x \cdot \frac{3}{2}(1 - x^2)\,dx$$

$$= \frac{3}{2}\int_{0}^{1}(x - x^3)\,dx = \frac{3}{2}\left(\frac{x^2}{2} - \frac{x^4}{4}\right)\Bigg|_{x=0}^{x=1} = \frac{3}{8}$$

■

Cuando la función de densidad de probabilidad $f(x)$ especifica un modelo para la distribución de valores en una población numérica, entonces μ es la media de la población, la cual es la medida más frecuentemente utilizada de la ubicación o centro de la población.

Con frecuencia se desea calcular el valor esperado de alguna función $h(X)$ de la variable aleatoria X. Si se piensa en $h(X)$ como una nueva variable aleatoria Y, se utilizan técnicas de estadística matemática para obtener la función de densidad de probabilidad de Y, y $E(Y)$ se calcula a partir de la definición. Afortunadamente, como en el caso discreto, existe una forma más fácil de calcular $E[h(X)]$.

PROPOSICIÓN

Si X es una variable aleatoria continua con función de densidad de probabilidad $f(x)$ y $h(X)$ es cualquier función de X, entonces

$$E[h(X)] = \mu_{h(X)} = \int_{-\infty}^{\infty} h(x) \cdot f(x)\,dx$$

Ejemplo 4.11 Dos especies compiten en una región por el control de una cantidad limitada de cierto recurso. Sea X = la proporción del recurso controlado por la especie 1 y suponga que la función de densidad de probabilidad de X es

$$f(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{de lo contrario} \end{cases}$$

la cual es una distribución uniforme en [0, 1]. (En su libro *Ecological Diversity*, E. C. Pielou llama a esto el modelo del "palo roto" para la asignación de recursos, puesto que es análogo a la ruptura de un palo en un lugar seleccionado al azar.) Entonces la especie que controla la mayor parte de este recurso controla la cantidad

$$h(X) = \text{máx}(X, 1 - X) = \begin{cases} 1 - X & \text{si } 0 \le X < \frac{1}{2} \\ X & \text{si } \frac{1}{2} \le X \le 1 \end{cases}$$

La cantidad esperada controlada por la especie que controla la mayor parte es entonces

$$E[h(X)] = \int_{-\infty}^{\infty} \text{máx}(x, 1 - x) \cdot f(x)\,dx = \int_{0}^{1} \text{máx}(x, 1 - x) \cdot 1\,dx$$

$$= \int_{0}^{1/2}(1 - x) \cdot 1\,dx + \int_{1/2}^{1} x \cdot 1\,dx = \frac{3}{4}$$

■

Para $h(X)$, una función lineal, $E[h(X)] = E(aX + b) = aE(X) + b$.

En el caso discreto, la varianza de X se definió como la desviación al cuadrado esperada con respecto a μ y se calculó por medio de suma. En este caso de nuevo la integración reemplaza a la suma.

DEFINICIÓN

La **varianza** de una variable aleatoria continua X con función de densidad de probabilidad $f(x)$ y valor medio μ es

$$\sigma_X^2 = V(X) = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x)dx = E[(X - \mu)^2]$$

La **desviación estándar** (DE) de X es $\sigma_X = \sqrt{V(X)}$.

La varianza y la desviación estándar dan medidas cuantitativas de cuánta dispersión hay en la distribución o población de valores x. Una vez más σ es aproximadamente del tamaño de una desviación típica de μ. El cálculo de σ^2 se facilita mediante el uso de la fórmula abreviada similar a la utilizada en el caso discreto.

PROPOSICIÓN

$$V(X) = E(X^2) - [E(X)]^2$$

Ejemplo 4.12 (Continuación del ejemplo 4.10)

Para X = ventas semanales de grava, se calculó $E(X) = \frac{3}{8}$. Como

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)\,dx = \int_0^1 x^2 \cdot \frac{3}{2}(1 - x^2)\,dx$$

$$= \int_0^1 \frac{3}{2}(x^2 - x^4)\,dx = \frac{1}{5}$$

$$V(X) = \frac{1}{5} - \left(\frac{3}{8}\right)^2 = \frac{19}{320} = .059 \text{ y } \sigma_X = .244$$ ■

Cuando $h(X) = aX + b$, el valor esperado y la varianza de $h(X)$ satisfacen las mismas propiedades que en el caso discreto: $E[h(X)] = a\mu + b$ y $V[h(X)] = a^2 \cdot \sigma^2$.

EJERCICIOS Sección 4.2 (11–27)

11. Sea X la cantidad de tiempo que un libro en reserva de dos horas está realmente prestado y supongamos que la función de distribución acumulativa es

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x^2}{4} & 0 \le x < 2 \\ 1 & 2 \le x \end{cases}$$

Use la función de distribución acumulativa para calcular lo siguiente:

a. $P(X \le 1)$
b. $P(.5 \le X \le 1)$
c. $P(X > 1.5)$
d. La mediana del tiempo de préstamo $\tilde{\mu}$ [resolver $.5 = F(\tilde{\mu})$]
e. $F'(x)$ para obtener la función de densidad $f(x)$
f. $E(X)$
g. $V(X)$ y σ_X
h. Si al prestatario se le cobra una cantidad $h(X = X^2$ cuando el tiempo de préstamo es X, calcule el cobro esperado $E[h(X)]$.

12. La función de distribución acumulativa de X (= error de medición) del ejercicio 3 es

$$F(x) = \begin{cases} 0 & x < -2 \\ \frac{1}{2} + \frac{3}{32}\left(4x - \frac{x^3}{3}\right) & -2 \le x < 2 \\ 1 & 2 \le x \end{cases}$$

a. Calcule $P(X < 0)$.
b. Calcule $P(-1 < X < 1)$.
c. Calcule $P(.5 < X)$.
d. Verifique que $f(x)$ es la dada en el ejercicio 3 obteniendo $F'(x)$.
e. Verifique que $\tilde{\mu} = 0$.

13. El ejemplo 4.5 introdujo el concepto de intervalo de tiempo en el flujo de tránsito y propuso una distribución particular para X = el intervalo de tiempo entre dos carros consecutivos seleccionados al azar (s). Suponga que en un entorno de tránsito diferente, la distribución del intervalo de tiempo tiene la forma

$$f(x) = \begin{cases} \dfrac{k}{x^4} & x > 1 \\ 0 & x \le 1 \end{cases}$$

a. Determine el valor de k con el cual $f(x)$ es una función de densidad de probabilidad legítima.
b. Obtenga la función de distribución acumulativa.
c. Use la función de distribución acumulativa de (b) para determinar la probabilidad de que el intervalo de tiempo exceda de 2 segundos y también la probabilidad de que el intervalo sea de entre 2 y 3 segundos.
d. Obtenga un valor medio del intervalo de tiempo y su desviación estándar.
e. ¿Cuál es la probabilidad de que el intervalo de tiempo quede dentro de 1 desviación estándar del valor medio?

14. El artículo "Modeling Sediment and Water Column Interactions for Hidrophobic Pollutants" (*Water Research*, 1984: 1169–1174) sugiere la distribución uniforme en el intervalo (7.5, 20) como modelo de profundidad (cm) de la capa de bioturbación en sedimento en una región.
a. ¿Cuáles son la media y la varianza de la profundidad?
b. ¿Cuál es la función de distribución acumulativa de la profundidad?
c. ¿Cuál es la probabilidad de que la profundidad observada sea cuando mucho de 10? ¿Entre 10 y 15?
d. ¿Cuál es la probabilidad de que la profundidad observada esté dentro de 1 desviación estándar del valor medio? ¿Dentro de 2 desviaciones estándar?

15. Sea X la cantidad de espacio ocupado por un artículo colocado en un contenedor de 1 pie^3. La función de densidad de probabilidad de X es

$$f(x) = \begin{cases} 90x^8(1 - x) & 0 < x < 1 \\ 0 & \text{de lo contrario} \end{cases}$$

a. Grafique la función de densidad de probabilidad. Luego obtenga la función de distribución acumulativa de X y grafíquela.
b. ¿Cuál es $P(X \le .5)$ [es decir, $F(.5)$]?
c. Con la función de distribución acumulativa de (a), ¿cuál es $P(.25 < X \le .5)$? ¿Cuál es $P(.25 \le X \le .5)$?
d. ¿Cuál es el 75avo percentil de la distribución?
e. Calcule $E(X)$ y σ_X.
f. ¿Cuál es la probabilidad de que X esté a más de 1 desviación estándar de su valor medio?

16. Responda los incisos (a)–(f) del ejercicio 15 con X = tiempo de disertación después de la hora dado en el ejercicio 5.

17. Si la distribución de X en el intervalo $[A, B]$ es uniforme
a. Obtenga una expresión para el $(100p)^{\text{o}}$ percentil.
b. Calcule $E(X)$, $V(X)$ y σ_X.
c. Con n, un entero positivo, calcule $E(X^n)$.

18. Sea X el voltaje a la salida de un micrófono y suponga que X tiene una distribución uniforme en el intervalo de -1 a 1. El voltaje es procesado por un "limitador duro" con valores de corte de $-.5$ y .5, de modo que la salida del limitador es una variable aleatoria Y relacionada con X por $Y = X$ si $|X| \le .5$, $Y = .5$ si $X > .5$ y $Y = -.5$ si $X < -.5$.
a. ¿Cuál es $P(Y = .5)$?
b. Obtenga la función de distribución acumulativa de Y y grafíquela.

19. Sea X una variable aleatoria continua con función de distribución acumulativa

$$F(x) = \begin{cases} 0 & x \le 0 \\ \dfrac{x}{4}\left[1 + \ln\left(\dfrac{4}{x}\right)\right] & 0 < x \le 4 \\ 1 & x > 4 \end{cases}$$

[Este tipo de función de distribución acumulativa es sugerido en el artículo "Variability in Measured Bedload-Transport Rates" (*Water Resources Bull.*, 1985: 39–48) como modelo de cierta variable hidrológica.] ¿Cuál es
a. $P(X \le 1)$?
b. $P(1 \le X \le 3)$?
c. La función de densidad de probabilidad de X?

20. Considere la función de densidad de probabilidad del tiempo de espera total Y de dos camiones

$$f(y) = \begin{cases} \dfrac{1}{25}y & 0 \le y < 5 \\ \dfrac{2}{5} - \dfrac{1}{25}y & 5 \le y \le 10 \\ 0 & \text{de lo contrario} \end{cases}$$

introducida en el ejercicio 8.
a. Calcule y trace la función de distribución acumulativa de Y. [*Sugerencia*: considere por separado $0 \le y < 5$ y $5 \le y \le 10$ al calcular $F(y)$. Una gráfica de la función de densidad de probabilidad debe ser útil.]
b. Obtenga una expresión para el $(100p)^{\text{o}}$ percentil. [*Sugerencia*: Considere por separado $0 < p < .5$ y $.5 < p < 1$.]
c. Calcule $E(Y)$ y $V(Y)$. ¿Cómo se comparan estos valores con el tiempo de espera probable y la varianza de un solo camión cuando el tiempo está uniformemente distribuido en [0, 5]?

21. Un ecólogo desea marcar una región de muestreo circular de 10 m de radio. Sin embargo, el radio de la región resultante en realidad es una variable aleatoria R con función de densidad de probabilidad

$$f(r) = \begin{cases} \dfrac{3}{4}[1 - (10 - r)^2] & 9 \le r \le 11 \\ 0 & \text{de lo contrario} \end{cases}$$

¿Cuál es el área esperada de la región circular resultante?

22. La demanda semanal de gas propano (en miles de galones) de una instalación particular es una variable aleatoria X con función de densidad de probabilidad

$$f(x) = \begin{cases} 2\left(1 - \dfrac{1}{x^2}\right) & 1 \le x \le 2 \\ 0 & \text{de lo contrario} \end{cases}$$

a. Calcule la función de distribución acumulativa de X.
b. Obtenga una expresión para el $(100p)^{\text{o}}$ percentil. ¿Cuál es el valor de $\tilde{\mu}$?
c. Calcule $E(X)$ y $V(X)$.
d. Si 1500 galones están en existencia al principio de la semana y no se espera ningún nuevo suministro durante la semana, ¿cuántos de los 1500 galones se espera que queden al final de la semana? [*Sugerencia*: sea $h(x)$ = cantidad que queda cuando la demanda = x.]

23. Si la temperatura a la cual cierto compuesto se funde es una variable aleatoria con valor medio de 120°C y desviación estándar de 2°C, ¿cuáles son la temperatura media y la desviación estándar medidas en °F? [*Sugerencia*: °F = 1.8°C + 32.]

24. La función de densidad de probabilidad de Pareto de X es

$$f(x; k, \theta) = \begin{cases} \dfrac{k \cdot \theta^k}{x^{k+1}} & x \geq \theta \\ 0 & x < \theta \end{cases}$$

introducida en el ejercicio 10.

a. Si $k > 1$, calcule $E(X)$.

b. ¿Qué se puede decir sobre $E(X)$ si $k = 1$?

c. Si $k > 2$, demuestre que $V(X) = k\theta^2(k - 1)^{-2}(k - 2)^{-1}$.

d. Si $k = 2$, ¿qué se puede decir sobre $V(X)$?

e. ¿Qué condiciones en cuanto a k son necesarias para garantizar que $E(X^n)$ es finito?

25. Sea X la temperatura en °C a la cual ocurre una reacción química y sea Y la temperatura en °F (así que $Y = 1.8X + 32$).

a. Si la mediana de la distribución X es $\tilde{\mu}$, demuestre que $1.8\tilde{\mu} + 32$ es la mediana de la distribución Y.

b. ¿Cómo está relacionado el 90º percentil de la distribución Y con el 90º percentil de la distribución X? Verifique su conjetura.

c. Más generalmente, si $Y = aX + b$, ¿cómo está relacionado cualquier percentil de la distribución Y con el percentil correspondiente de la distribución X?

26. Sea X los gastos médicos totales (en miles de dólares) en que incurre un individuo particular durante un año dado. Aunque X es una variable aleatoria discreta, suponga que su distribución es bastante bien aproximada por una distribución continua con función de densidad de probabilidad $f(x) = k(1 + x/2.5)^{-7}$ para $x \geq 0$.

a. ¿Cuál es el valor de k?

b. Grafique la función de densidad de probabilidad de X.

c. ¿Cuáles son el valor esperado y la desviación estándar de los gastos médicos totales?

d. Este individuo está cubierto por un plan de aseguramiento que le impone una provisión deducible de $500 (así que los primeros $500 de gastos son pagados por el individuo). Luego el plan pagará 80% de cualquier gasto adicional que exceda de $500 y el pago máximo por parte del individuo (incluida la cantidad deducible) es de $2500. Sea Y la cantidad de gastos médicos de este individuo pagados por la compañía de seguros. ¿Cuál es el valor esperado de Y? [*Sugerencia*: primero indague qué valor de X corresponde al gasto máximo que sale del bolsillo de $2500. Luego escriba una expresión para Y como una función de X (la cual implique varios precios diferentes) y calcule el valor esperado de la función.]

27. Cuando se lanza un dardo a un blanco circular, considere la ubicación del punto de aterrizaje con respecto al centro del blanco. Sea X el ángulo en grados medido con respecto a la horizontal y suponga que X está uniformemente distribuida en [0, 360]. Defina Y como la variable transformada $Y = h(X) = (2\pi/360)X - \pi$, por lo tanto, Y es el ángulo medido en radianes y Y está entre $-\pi$ y π. Obtenga $E(Y)$ y σ_Y obteniendo primero $E(X)$ y σ_X y luego utilizando el hecho de que $h(X)$ es una función lineal de X.

4.3 Distribución normal

La **distribución normal** es la más importante en toda la probabilidad y estadística. Muchas poblaciones numéricas tienen distribuciones que pueden ser representadas muy fielmente por una curva normal apropiada. Los ejemplos incluyen estaturas, pesos y otras características físicas (el famoso artículo *Biometrika* 1903 "On the Laws of Inheritance in Man" discutió muchos ejemplos de esta clase), errores de medición en experimentos científicos, mediciones antropométricas en fósiles, tiempos de reacción en experimentos psicológicos, mediciones de inteligencia y aptitud, calificaciones en varios exámenes y numerosas medidas e indicadores económicos. Además, aun cuando las variables individuales no estén normalmente distribuidas, las sumas y promedios de las variables en condiciones adecuadas tendrán de manera aproximada una distribución normal; éste es el contenido del teorema del límite central discutido en el siguiente capítulo.

DEFINICIÓN

Se dice que una variable aleatoria continua X tiene una **distribución normal** con parámetros μ y σ (o μ y σ^2), donde $-\infty < \mu < \infty$ y $0 < \sigma$, si la función de densidad de probabilidad de X es

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)} \quad -\infty < x < \infty \qquad (4.3)$$

De nuevo e denota la base del sistema de logaritmos naturales y es aproximadamente igual a 2.71828 y π representa la conocida constante matemática con un valor aproximado de 3.14159. El enunciado de que X está normalmente distribuida con los parámetros μ y σ^2 a menudo se abrevia como $X \sim N(\mu, \sigma^2)$.

Claramente $f(x; \mu, \sigma) \geq 0$ aunque se tiene que utilizar un argumento de cálculo un tanto complicado para verificar que $\int_{-\infty}^{\infty} f(x; \mu, \sigma)\, dx = 1$. Se puede demostrar que $E(X) = \mu$ y $V(X) = \sigma^2$, de modo que los parámetros son la media y la desviación estándar de X. La figura 4.13 representa gráficas de $f(x; \mu, \sigma)$ de varios pares diferentes (μ, σ). Cada curva de densidad es simétrica con respecto a μ y acampanada, de modo que el centro de la campana (punto de simetría) es tanto la media de la distribución como la mediana. El valor de σ es la distancia desde μ hasta los puntos de inflexión de la curva (los puntos donde la curva cambia de dirección hacia abajo o hacia arriba). Los grandes valores de σ producen gráficas que están bastante extendidas en torno a μ, en tanto que los valores pequeños de σ dan gráficas con una alta cresta sobre μ y la mayor parte del área bajo la gráfica bastante cerca de μ. Así pues, una σ grande implica que se puede observar muy bien un valor de X alejado de μ, en tanto que dicho valor es bastante improbable cuando σ es pequeña.

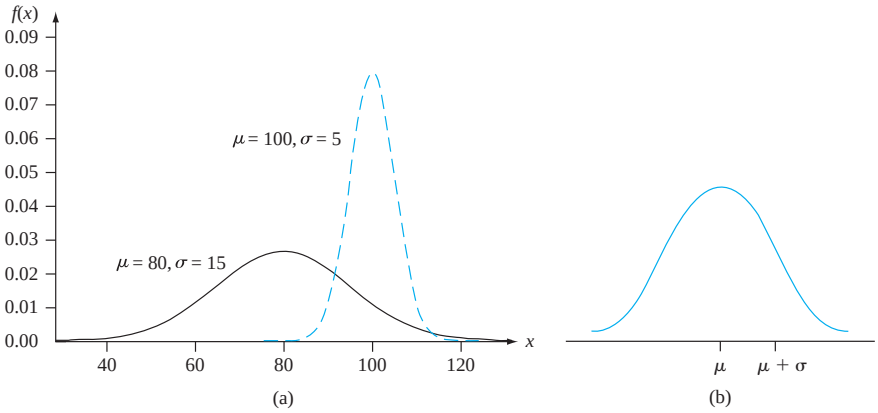


Figura 4.13 (a) Dos curvas diferentes de densidad normal (b) Visualización de μ y σ para una distribución normal

Distribución normal estándar

El cálculo de $P(a \leq X \leq b)$ cuando X es una variable aleatoria normal con parámetros μ y σ, requiere determinar

$$\int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}\, dx \tag{4.4}$$

Ninguna de las técnicas estándar de integración puede ser utilizada para lograr esto. En cambio, con $\mu = 0$ y $\sigma = 1$, se calculó la expresión (4.4) por medio de técnicas numéricas y se tabuló para ciertos valores de a y b. Esta tabla también puede ser utilizada para calcular probabilidades con cualesquiera otros valores de μ y σ considerados.

DEFINICIÓN

La distribución normal con valores de parámetro $\mu = 0$ y $\sigma = 1$ se llama **distribución normal estándar.** Una variable aleatoria que tiene una distribución normal estándar se llama **variable aleatoria normal estándar** y se denotará por Z. La función de densidad de probabilidad de Z es

$$f(z; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad -\infty < z < \infty$$

La gráfica de $f(z; 0, 1)$ se llama curva *normal estándar* (o z). Sus puntos de inflexión estan en 1 y -1. La función de distribución acumulativa de Z es $P(Z \le z) = \int_{-\infty}^{z} f(y; 0, 1)\, dy$ la cual será denotada por $\Phi(z)$.

La distribución normal estándar no siempre sirve como modelo de una población que surge naturalmente. En cambio, es una distribución de referencia de la que se puede obtener información sobre otra distribución normal. La tabla A.3 del apéndice, da $\Phi(z) = P(Z \le z)$, el área bajo la curva de densidad normal estándar a la izquierda de z con $z = -3.49, -3.48, \ldots, 3.48, 3.49$. La figura 4.14 ilustra el tipo de área acumulativa (probabilidad) tabulada en la tabla A.3. Con esta tabla, varias otras probabilidades que implican Z pueden ser calculadas.

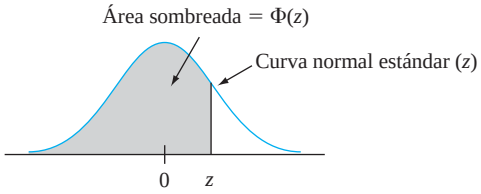


Figura 4.14 Áreas acumulativas normales estándar tabuladas en la tabla A.3 del apéndice

Ejemplo 4.13 Determínense las siguientes probabilidades normales estándar: (a) $P(Z \le 1.25)$, (b) $P(Z > 1.25)$, (c) $P(Z \le -1.25)$ y (d) $P(-.38 \le Z \le 1.25)$.

a. $P(Z \le 1.25) = \Phi(1.25)$, una probabilidad tabulada en la tabla A.3 del apéndice en la intersección de la fila 1.2 y la columna .05. El número allí es .8944, así que $P(Z \le 1.25) = .8944$. La figura 4.15(a) ilustra esta probabilidad.

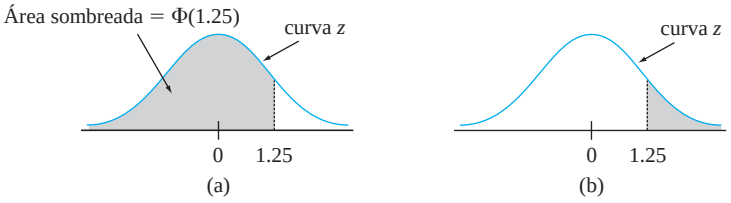


Figura 4.15 Áreas (probabilidades) de curvas normales del ejemplo 4.13

b. $P(Z > 1.25) = 1 - P(Z \le 1.25) = 1 - \Phi(1.25)$, el área bajo la curva z a la derecha de 1.25 (un área de cola superior). En ese caso $\Phi(1.25) = .8944$ implica que $P(Z > 1.25) = .1056$. Como Z es una variable aleatoria continua, $P(Z \ge 1.25) = .1056$. Véase la figura 4.15(b).

c. $P(Z \le -1.25) = \Phi(-1.25)$, un área de cola inferior. Directamente de la tabla A.3 del apéndice, $\Phi(-1.25) = .1056$. Por simetría de la curva z, ésta es la misma respuesta del inciso (b).

d. $P(-.38 \le Z \le 1.25)$ es el área bajo la curva normal estándar sobre el intervalo cuyo punto extremo izquierdo es $-.38$ y cuyo punto extremo derecho es 1.25. Según la sección 4.2, si X es una variable aleatoria continua con función de distribución acumulativa $F(x)$, entonces $P(a \le X \le b) = F(b) - F(a)$. Por lo tanto $P(-.38 \le Z \le 1.25) = \Phi(1.25) - \Phi(-.38) = .8944 - .3520 = .5424$. (Véase la figura 4.16.)

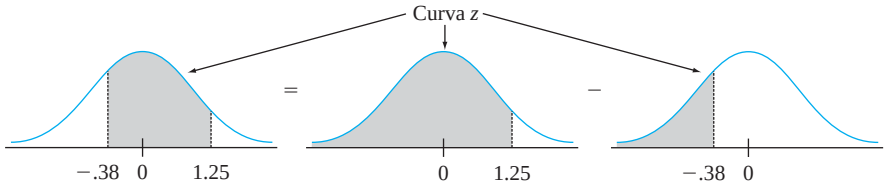


Figura 4.16 $P(-.38 \le Z \le 1.25)$ como la diferencia entre dos áreas acumulativas ■

Percentiles de la distribución normal estándar

Con cualquier p entre 0 y 1, se puede utilizar la tabla A.3 del apéndice para obtener el $(100p)^{o}$ percentil de la distribución normal estándar.

Ejemplo 4.14 El 99º percentil de la distribución normal estándar es ese valor sobre el eje horizontal tal que el área bajo la curva z a la izquierda de dicho valor es .9900. La tabla A.3 del apéndice da con z fija el área bajo la curva normal estándar a la izquierda de z, mientras que aquí se tiene el área y se desea el valor de z. Éste es el problema "inverso" a $P(Z \le z) = ?$ así que la tabla se utiliza a la inversa: encuentre en la mitad de la tabla .9900; la fila y la columna en la que se encuentra identifica el 99° percentil z. En este caso .9901 queda en la intersección de la fila 2.3 y la columna .03, así que el 99° percentil es (aproximadamente) $z = 2.33$. (Véase la figura 4.17). Por simetría, el primer percentil está tan debajo de 0 como el 99° está sobre 0, así que es igual a -2.33 (1% queda debajo del primero y también sobre el 99°). (Véase la figura 4.18.)

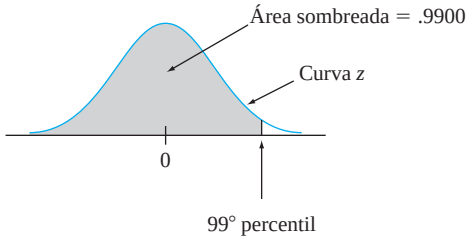


Figura 4.17 Localización del 99º percentil

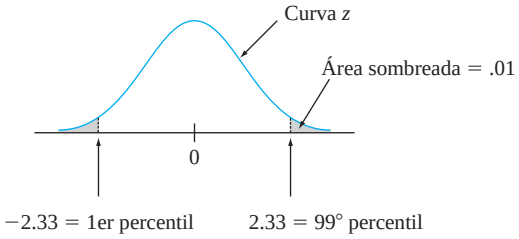


Figura 4.18 Relación entre el 1º y el 99º percentiles ■

En general, la fila y la columna de la tabla A.3 del apéndice, donde la entrada p está localizado identifican el $(100p)^{o}$ percentil (p. ej., el 67º percentil se obtiene localizando .6700 en el cuerpo de la tabla, la cual da $z = .44$). Si p no aparece, a menudo se utiliza el número más cercano a él, aunque la interpolación lineal da una respuesta más precisa. Por ejemplo, para encontrar el 95º percentil, se busca .9500 adentro de la tabla. Aunque .9500 no aparece, tanto .9495 como .9505 sí, correspondientes a $z = 1.64$ y 1.65, respectivamente.

Como .9500 está a la mitad entre las dos probabilidades que sí aparecen, se utilizará 1.645 como el 95º percentil y -1.645 como el 5º percentil.

Notación z_α para valores z críticos

En inferencia estadística se necesitan valores sobre el eje horizontal z que capturen ciertas áreas de cola pequeña bajo la curva normal estándar.

> **Notación**
>
> z_α denotará el valor sobre el eje z para el cual α del área bajo la curva z queda a la derecha de z_α. (Véase la figura 4.19.)

Por ejemplo, $z_{.10}$ captura el área de cola superior .10, y $z_{.01}$ captura el área de cola superior .01.

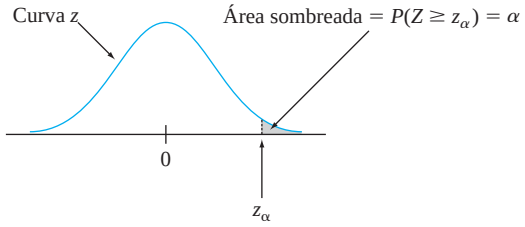


Figura 4.19 Notación z_α ilustrada

Como α del área bajo la curva z queda a la derecha de z_α, $1 - \alpha$ del área queda a su izquierda. Por lo tanto, z_α *es el* $100(1 - \alpha)$º *percentil de la distribución normal estándar*. Por simetría el área bajo la curva normal estándar a la izquierda de $-z_\alpha$ también es α. Los valores z_α en general se conocen como **valores críticos z.** La tabla 4.1 incluye los percentiles z y los valores z_α más útiles.

Tabla 4.1 Percentiles de la distribución normal estándar y valores críticos

Percentil	90	95	97.5	99	99.5	99.9	99.95
α (área de la cola)	.1	.05	.025	.01	.005	.001	.0005
$z_\alpha = 100(1 - \alpha)$º percentil	1.28	1.645	1.96	2.33	2.58	3.08	3.27

Ejemplo 4.15 $z_{.05}$ es el $100(1 - .05)^\circ = 95^\circ$ percentil de la distribución normal estándar, por lo tanto $z_{.05} = 1.645$. El área bajo la curva normal estándar a la izquierda de $-z_{.05}$ también es .05. (Véase la figura 4.20.)

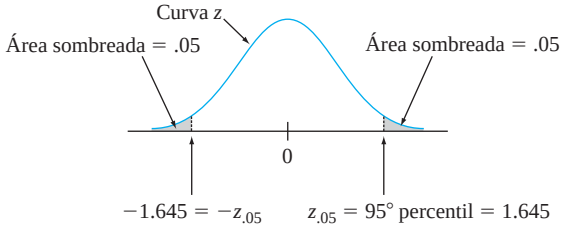


Figura 4.20 Determinación de $z_{.05}$ ■

Distribuciones normales no estándar

Cuando $X \sim N(\mu, \sigma^2)$, las probabilidades que implican X se calculan "estandarizando". La **variable estandarizada** es $(X - \mu)/\sigma$. Al restar μ la media cambia de μ a cero y luego al dividir entre σ cambian las escalas de la variable de modo que la desviación estándar es 1 en lugar de σ.

PROPOSICIÓN

Si X tiene una distribución normal con media μ y desviación estándar σ, entonces

$$Z = \frac{X - \mu}{\sigma}$$

tiene una distribución normal estándar. Por lo tanto

$$P(a \le X \le b) = P\left(\frac{a - \mu}{\sigma} \le Z \le \frac{b - \mu}{\sigma}\right)$$

$$= \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

$$P(X \le a) = \Phi\left(\frac{a - \mu}{\sigma}\right) \qquad P(X \ge b) = 1 - \Phi\left(\frac{b - \mu}{\sigma}\right)$$

La idea clave de la proposición es que estandarizando, cualquier probabilidad que implique X puede ser expresada como una probabilidad que implica una variable aleatoria normal estándar Z, de modo que se pueda utilizar la tabla A.3 del apéndice. Esto se ilustra en la figura 4.21. La proposición se comprueba escribiendo la función de distribución acumulativa de $Z = (X - \mu)/\sigma$ como

$$P(Z \le z) = P(X \le \sigma z + \mu) = \int_{-\infty}^{\sigma z + \mu} f(x; \mu, \sigma)dx$$

Utilizando un resultado del cálculo, esta integral puede ser diferenciada con respecto a z para que dé la función de densidad de probabilidad deseada $f(z; 0, 1)$.

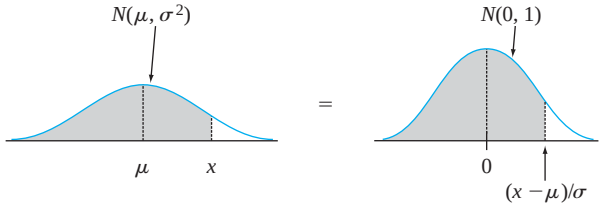


Figura 4.21 Igualdad de áreas de curvas normales estándar y no estándar

Ejemplo 4.16 El tiempo que requiere un conductor para reaccionar a las luces de freno de un vehículo que está desacelerando es crítico para evitar colisiones por alcance. El artículo "Fast-Rise Brake Lamp as a Collision-Prevention Device" (*Ergonomics*, 1993: 391–395), sugiere que el tiempo de reacción de respuesta en tráfico a una señal de luces de freno estándar puede ser modelado con una distribución normal que tiene un valor medio de 1.25 s y desviación

estándar de .46 s. ¿Cuál es la probabilidad de que el tiempo de reacción sea de entre 1.00 y 1.75 segundos? Si X denota el tiempo de reacción, entonces estandarizando se obtiene

$$1.00 \le X \le 1.75$$

si y sólo si

$$\frac{1.00 - 1.25}{.46} \le \frac{X - 1.25}{.46} \le \frac{1.75 - 1.25}{.46}$$

Por lo tanto

$$\begin{aligned} P(1.00 \le X \le 1.75) &= P\left(\frac{1.00 - 1.25}{.46} \le Z \le \frac{1.75 - 1.25}{.46}\right) \\ &= P(-.54 \le Z \le 1.09) = \Phi(1.09) - \Phi(-.54) \\ &= .8621 - .2946 = .5675 \end{aligned}$$

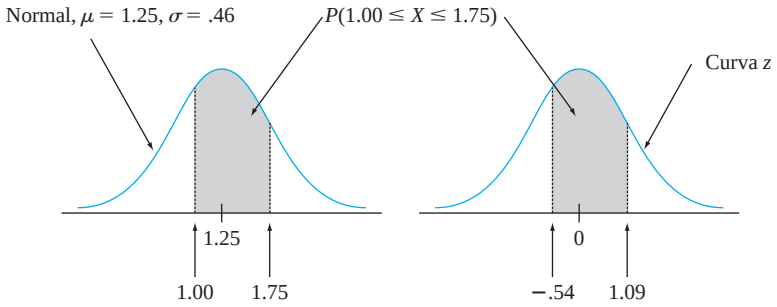


Figura 4.22 Curvas normales del ejemplo 4.16

Esto se ilustra en la figura 4.22. Asimismo, si se ven los 2 segundos como un tiempo de reacción críticamente largo, la probabilidad de que el tiempo de reacción real exceda este valor es

$$P(X > 2) = P\left(Z > \frac{2 - 1.25}{.46}\right) = P(Z > 1.63) = 1 - \Phi(1.63) = .0516$$ ■

Estandarizar cantidades no lleva a nada más que a calcular una distancia del valor medio y luego reexpresarla como algún número de desviaciones estándar. Por lo tanto, si $\mu = 100$ y $\sigma = 15$, entonces $x = 130$ corresponde a $z = (130 - 100)/15 = 30/15 = 2.00$. Es decir, 130 está a 2 desviaciones estándar sobre (a la derecha de) el valor medio. Asimismo, estandarizando 85 se obtiene $(85 - 100)/15 = -1.00$, por lo tanto 85 está a 1 desviación estándar por debajo de la media. La tabla z se aplica a *cualquier* distribución normal siempre que se piense en función del número de desviaciones estándar con respecto al valor medio.

Ejemplo 4.17 Se sabe que el voltaje de ruptura de un diodo seleccionado al azar de un tipo particular está normalmente distribuido. ¿Cuál es la probabilidad de que el voltaje de ruptura de un diodo esté dentro de 1 desviación estándar de su valor medio? Esta pregunta puede ser respondida sin conocer μ o σ, en tanto se sepa que la distribución es normal; la respuesta es la misma para *cualquier* distribución normal:

$$\begin{aligned} P(X \text{ está dentro de 1 desviación estándar de su media}) &= P(\mu - \sigma \le X \le \mu + \sigma) \\ &= P\left(\frac{\mu - \sigma - \mu}{\sigma} \le Z \le \frac{\mu + \sigma - \mu}{\sigma}\right) \\ &= P(-1.00 \le Z \le 1.00) \\ &= \Phi(1.00) - \Phi(-1.00) = .6826 \end{aligned}$$

La probabilidad de que X esté dentro de 2 desviaciones estándar de su media es $P(-2.00 \le Z \le 2.00) = .9544$ y dentro de 3 desviaciones estándar de su media es $P(-3.00 \le Z \le 3.00) = .9974$. ■

Los resultados del ejemplo 4.17 a menudo se reportan en forma de porcentaje y se les conoce como la *regla empírica* (porque la evidencia empírica ha demostrado que los histogramas de datos reales con frecuencia pueden ser aproximados por curvas normales).

Si la distribución de la población de una variable es (aproximadamente) normal, entonces

1. Aproximadamente 68% de los valores están dentro de 1 DE de la media.
2. Aproximadamente 95% de los valores están dentro de 2 DE de la media.
3. Aproximadamente 99.7% de los valores están dentro de 3 DE de la media.

En realidad es inusual observar un valor de una población normal que esté mucho más lejos de 2 desviaciones estándar de μ. Estos resultados serán importantes en el desarrollo de procedimientos de prueba de hipótesis en capítulos posteriores.

Percentiles de una distribución normal arbitraria

El $(100p)^{\circ}$ percentil de una distribución normal con media μ y desviación estándar σ es fácil de relacionar con el $(100p)^{\circ}$ percentil de la distribución normal estándar.

PROPOSICIÓN

$$\begin{pmatrix}(100p)^{\circ}\text{ percentil}\\ \text{para }(\mu,\sigma)\text{ normal}\end{pmatrix} = \mu + \begin{bmatrix}(100p)^{\circ}\text{ para}\\ \text{normal estándar}\end{bmatrix} \cdot \sigma$$

Otra forma de decir esto es que si z es el percentil deseado de la distribución normal estándar, entonces el percentil deseado de la distribución (μ, σ) normal está a z desviaciones estándar de μ.

Ejemplo 4.18 La cantidad de agua destilada despachada por cierta máquina está normalmente distribuida con valor medio de 64 oz y desviación estándar de .78 oz. ¿Qué tamaño de contenedor c asegurará que ocurra rebosamiento sólo .5% del tiempo? Si X denota la cantidad despachada, la condición deseada es que $P(X > c) = .005$ o, en forma equivalente, que $P(X \le c) = .995$. Por lo tanto c es el 99.5º percentil de la distribución normal con $\mu = 64$ y $\sigma = .78$. El 99.5º percentil de la distribución normal estándar es 2.58, por lo tanto

$$c = \eta(.995) = 64 + (2.58)(.78) = 64 + 2.0 = 66 \text{ oz}$$

Esto se ilustra en la figura 4.23.

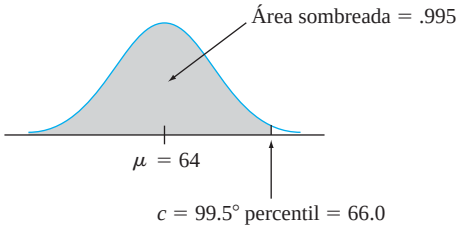


Figura 4.23 Distribución de la cantidad despachada en el ejemplo 4.18 ■

Distribución normal y poblaciones discretas

La distribución normal a menudo se utiliza como una aproximación a la distribución de valores en una población discreta. En semejantes situaciones se debe tener cuidado especial para asegurarse de que las probabilidades se calculen con precisión.

Ejemplo 4.19 Se sabe que el coeficiente intelectual en una población particular (medido con una prueba estándar) está más o menos normalmente distribuido con $\mu = 100$ y $\sigma = 15$. ¿Cuál es la probabilidad de que un individuo seleccionado al azar tenga un CI de por lo menos 125? Con $X =$ el CI de una persona seleccionada al azar, se desea $P(X \geq 125)$. La tentación en este caso es estandarizar $X \geq 125$ como en ejemplos previos. Sin embargo, la distribución de la población de coeficientes intelectuales en realidad es discreta, puesto que los coeficientes intelectuales son valores enteros. Así que la curva normal es una aproximación a un histograma de probabilidad discreto como se ilustra en la figura 4.24.

Los rectángulos del histograma están *centrados* en enteros, por lo que los coeficientes intelectuales de por lo menos 125 corresponden a rectángulos que comienzan en 124.5, la zona sombreada en la figura 4.24. Por lo tanto en realidad se desea el área bajo la curva aproximadamente normal a la derecha de 124.5. Si se estandariza este valor se obtiene $P(Z \geq 1.63) = .0516$, en tanto que si se estandariza 125 se obtiene $P(Z \geq 1.67) = .0475$. La diferencia no es grande, pero la respuesta .0516 es más precisa. Asimismo, $P(X = 125)$ sería aproximada por el área entre 124.5 y 125.5, puesto que el área bajo la curva normal sobre el valor único de 125 es cero.

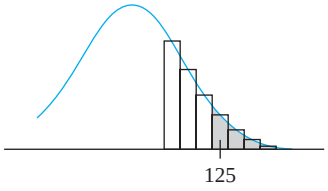


Figura 4.24 Aproximación normal a una distribución discreta ■

La corrección en cuanto a discrecionalidad de la distribución subyacente en el ejemplo 4.19 a menudo se llama **corrección de continuidad.** Es útil en la siguiente aplicación de la distribución normal al cálculo de probabilidades binomiales.

Aproximación de la distribución binomial

Recuérdese que el valor medio y la desviación estándar de una variable aleatoria binomial X son $\mu_X = np$ y $\sigma_X = \sqrt{npq}$, respectivamente. La figura 4.25 muestra un histograma de probabilidad binomial de la distribución binomial con $n = 20$, $p = .6$ con el cual $\mu = 20(.6) = 12$ y $\sigma = \sqrt{20(.6)(.4)} = 2.19$. Sobre el histograma de probabilidad se superpuso una curva normal con estas μ y σ. Aunque el histograma de probabilidad es un poco asimétrico (debido a que $p \neq .5$), la curva normal da una muy buena aproximación, sobre todo en la parte media de la figura. El área de cualquier rectángulo (probabilidad de cualquier valor X particular), excepto las de los localizados en las colas extremas, puede ser aproximada con precisión mediante el área de la curva normal correspondiente. Por ejemplo, $P(X = 10) = B(10; 20, .6) - B(9; 20, .6) = .117$, mientras que el área bajo la curva normal entre 9.5 y 10.5 es $P(-1.14 \leq Z \leq -.68) = .1212$.

Más generalmente, en tanto que el histograma de probabilidad binomial no sea demasiado asimétrico, las probabilidades binomiales pueden ser aproximadas muy bien por áreas de curva normal. Se acostumbra entonces decir que X tiene aproximadamente una distribución normal.

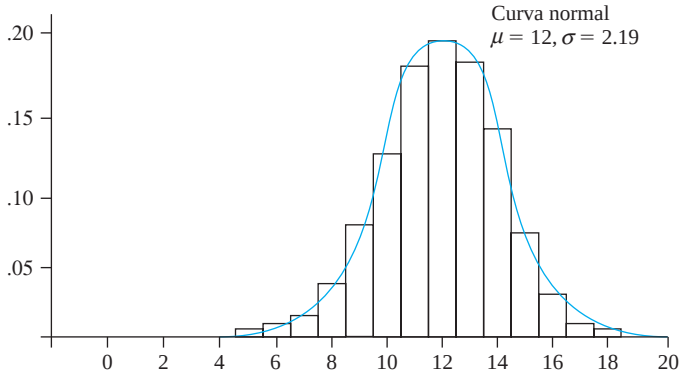


Figura 4.25 Histograma de probabilidad binomial para $n = 20$, $p = .6$ con curva de aproximación normal sobrepuesta

PROPOSICIÓN

Sea X una variable aleatoria binomial basada en n ensayos con probabilidad de éxito p. Entonces, si el histograma de probabilidad binomial no es demasiado asimétrico, X tiene aproximadamente una distribución normal con $\mu = np$ y $\sigma = \sqrt{npq}$. En particular, con $x =$ un valor posible de X,

$$P(X \le x) = B(x, n, p) \approx \begin{pmatrix}\text{área bajo la curva normal} \\ \text{a la izquierda de } x + .5\end{pmatrix}$$
$$= \Phi\left(\frac{x + .5 - np}{\sqrt{npq}}\right)$$

En la práctica, la aproximación es adecuada siempre que $np \ge 10$ y $nq \ge 10$, puesto que en ese caso existe bastante simetría en la distribución binomial subyacente.

Una comprobación directa de este resultado es bastante difícil. En el siguiente capítulo se verá que es una consecuencia de un resultado más general llamado teorema del límite central. Con toda honestidad, esta aproximación no es tan importante en el cálculo de probabilidad como una vez lo fue. Esto se debe a que los programas de computadora ahora son capaces de calcular probabilidades binomiales con exactitud para valores bastante grandes de n.

Ejemplo 4.20 Suponga que 25% de todos los estudiantes en una gran universidad pública reciben ayuda financiera. Sea X el número de estudiantes que reciben esta ayuda en una muestra aleatoria de tamaño 50, de modo que $p = .25$. Entonces $\mu = 12.5$ y $\sigma = 3.06$. Como $np = 50(.25) = 12.5 \ge 10$ y $nq = 37.5 \ge 10$, la aproximación puede ser aplicada con seguridad. La probablidad de que a lo más 10 estudiantes reciban ayuda es

$$P(X \le 10) = B(10; 50, .25) \approx \Phi\left(\frac{10 + .5 - 12.5}{3.06}\right)$$
$$= \Phi(-.65) = .2578$$

Asimismo, la probabilidad de que entre 5 y 15 (inclusive) de los estudiantes seleccionados reciban ayuda es

$$P(5 \le X \le 15) = B(15; 50, .25) - B(4; 50, .25)$$
$$\approx \Phi\left(\frac{15.5 - 12.5}{3.06}\right) - \Phi\left(\frac{4.5 - 12.5}{3.06}\right) = .8320$$

Las probabilidades exactas son .2622 y .8348, respectivamente, así que las aproximaciones son bastante buenas. En el último cálculo, la probabilidad $P(5 \le X \le 15)$ está siendo aproximada por el área bajo la curva normal entre 4.5 y 15.5; se utiliza la corrección de continuidad tanto para el límite superior como para el inferior. ■

Cuando el objetivo de la investigación es hacer una inferencia sobre una proporción de población p, el interés se enfocará en la proporción muestral de X/n éxitos y no en X. Como esta proporción es exactamente X multiplicada por la constante $1/n$, también tendrá aproximadamente una distribución normal (con media $\mu = p$ y desviación estándar $\sigma = \sqrt{pq/n}$, siempre que $np \ge 10$ y $nq \ge 10$. Esta aproximación normal es la base de varios procedimientos inferenciales que se discutirán en capítulos posteriores.

EJERCICIOS Sección 4.3 (28–58)

28. Sea Z una variable aleatoria normal estándar y calcule las siguientes probabilidades, trace las figuras siempre que sea apropiado.

a. $P(0 \le Z \le 2.17)$ **b.** $P(0 \le Z \le 1)$
c. $P(-2.50 \le Z \le 0)$ **d.** $P(-2.50 \le Z \le 2.50)$
e. $P(Z \le 1.37)$ **f.** $P(-1.75 \le Z)$
g. $P(-1.50 \le Z \le 2.00)$ **h.** $P(1.37 \le Z \le 2.50)$
i. $P(1.50 \le Z)$ **j.** $P(|Z| \le 2.50)$

29. En cada caso, determine el valor de la constante c que hace que el enunciado de probabilidad sea correcto.

a. $\Phi(c) = .9838$ **b.** $P(0 \le Z \le c) = .291$
c. $P(c \le Z) = .121$ **d.** $P(-c \le Z \le c) = .668$
e. $P(c \le |Z|) = .016$

30. Encuentre los siguientes percentiles de la distribución normal estándar. Interpole en los casos en que sea apropiado.

a. 91º **b.** 9º **c.** 75º
d. 25º **e.** 6º

31. Determine z_α para lo siguiente

a. $\alpha = .0055$ **b.** $\alpha = .09$
c. $\alpha = .663$

32. Suponga que la fuerza que actúa en una columna que ayuda a soportar un edificio es una variable aleatoria X normalmente distribuida con media de 15.0 kips y desviación estándar de 1.25 kips. Calcule las siguientes probabilidades por estandarización y luego use la tabla A.3

a. $P(X \le 15)$ **b.** $P(X \le 17.5)$
c. $P(X \ge 10)$ **d.** $P(14 \le X \le 18)$
e. $P(|X - 15| \le 3)$

33. Las mopeds (motos pequeñas con una cilindrada inferior a 50 cm^3) son muy populares en Europa debido a su movilidad, facilidad de uso y bajo costo. El artículo "Procedure to Verify the Maximum Speed of Automatic Transmission Mopeds in Periodic Motor Vehicle Inspections" (*J. of Automobile Engr.*, 2008: 1615–1623) describió un banco de pruebas rodante para determinar la velocidad máxima del vehículo. Se propone una distribución normal con valor medio de 46.8 km/h y desviación estándar de 1.75 km/h. Considere la posibilidad de seleccionar al azar una sola de esas mopeds.

a. ¿Cuál es la probabilidad de que la velocidad máxima sea a lo sumo 50 km/h?
b. ¿Cuál es la probabilidad de que la velocidad máxima sea al menos de 48 km/h?
c. ¿Cuál es la probabilidad de que la velocidad máxima difiera del valor medio por más de 1.5 desviaciones estándar?

34. El artículo "Reliability of Domestic-Waste Biofilm Reactors" (*J. of Envir. Engr.*, 1995: 785–790) sugiere que la concentración de sustrato (mg/cm^3) del afluente que llega a un reactor está normalmente distribuida con $\mu = .30$ y $\sigma = .06$.

a. ¿Cuál es la probabilidad de que la concentración exceda de .25?
b. ¿Cuál es la probabilidad de que la concentración sea cuando mucho de .10?
c. ¿Cómo caracterizaría el 5% más grande de todos los valores de concentración?

35. Suponga que el diámetro a la altura del pecho (pulg) de árboles de un tipo está normalmente distribuido con $\mu = 8.8$ y $\sigma = 2.8$, como se sugiere en el artículo "Simulating a Harvester-Forwarder Softwood Thinning" (*Forest Products J.*, mayo de 1997; 36–41).

a. ¿Cuál es la probabilidad de que el diámetro de un árbol seleccionado al azar será por lo menos de 10 pulg? ¿Y que exceda de 10 pulg?
b. ¿Cuál es la probabilidad de que el diámetro de un árbol seleccionado al azar sea de más de 20 pulg?
c. ¿Cuál es la probabilidad de que el diámetro de un árbol seleccionado al azar sea de entre 5 y 10 pulg?
d. ¿Qué valor c es tal que el intervalo $(8.8 - c, 8.8 + c)$ incluya 98% de todos los valores de diámetro?
e. Si se seleccionan cuatro árboles al azar, ¿cuál es la probabilidad de que por lo menos uno tenga un diámetro de más de 10 pulg?

36. La dispersión de las atomizaciones de pesticidas es una preocupación constante de los fumigadores y productores agrícolas.

La relación inversa entre el tamaño de gota y el potencial de deriva es bien conocida. El artículo "Effects of 2,4-D Formulation and Quinclorac on Spray Droplet Size and Deposition" (*Weed Technology*, 2005: 1030–1036) investigó los efectos de formulaciones de herbicidas en atomizaciones. Una figura en el artículo sugirió que la distribución normal con media de 1050 μm y desviación estándar de 150 μm fue un modelo razonable de tamaño de gotas de agua (el "tratamiento de control") pulverizada a través de una boquilla de 760 ml/min.

a. ¿Cuál es la probabilidad de que el tamaño de una sola gota sea de menos de 1500 μm? ¿Por lo menos de 1000 μm?

b. ¿Cuál es la probabilidad de que el tamaño de una sola gota sea de entre 1000 y 1500 μm?

c. ¿Cómo caracterizaría el 2% más pequeño de todas las gotas?

d. Si se miden los tamaños de cinco gotas independientemente seleccionadas, ¿cuál es la probabilidad de que por lo menos una exceda de 1500 μm?

37. Suponga que la concentración de cloruro en sangre (mmol/L) tiene una distribución normal con media de 104 y desviación estándar de 5 (información en el artículo "Mathematical Model of Chloride Concentration in Human Blood", *J. of Med. Engr. and Tech.*, 2006; 25–30, incluida una gráfica de probabilidad normal como se describe en la sección 4.6, apoya esta suposición).

a. ¿Cuál es la probabilidad de que la concentración de cloruro sea igual a 105? ¿Sea menor que 105? ¿Sea cuando mucho de 105?

b. ¿Cuál es la probabilidad de que la concentración de cloruro difiera de la media por más de 1 desviación estándar? ¿Depende esta probabilidad de los valores de μ y σ?

c. ¿Cómo caracterizaría el .1% más extremo de los valores de concentración de cloruro?

38. Hay dos máquinas disponibles para cortar corchos para usarse en botellas de vino. La primera produce corchos con diámetros que están normalmente distribuidos con media de 3 cm y desviación estándar de .1 cm. La segunda máquina produce corchos con diámetros que tienen una distribución normal con media de 3.04 cm y desviación estándar de .02 cm. Los corchos aceptables tienen diámetros de entre 2.9 y 3.1 cm. ¿Cuál máquina es más probable que produzca un corcho aceptable?

39. **a.** Si una distribución normal tiene $\mu = 30$ y $\sigma = 5$, ¿cuál es el 91º percentil de la distribución?

b. ¿Cuál es el 6 percentil de la distribución?

c. El ancho de una línea grabada en un "chip" de circuito integrado normalmente está distribuido con media de 3.000 μm y desviación estándar de .140. ¿Qué valor de ancho separa el 10% de las líneas más anchas del 90% restante?

40. El artículo "Monte Carlo Simulation–Tool for Better Understanding of LRFD" (*J. of Structural Engr.*, 1993: 1586–1599) sugiere que la resistencia a ceder (kg/pulg^2) de un acero grado A36 normalmente está distribuida con $\mu = 43$ y $\sigma = 4.5$.

a. ¿Cuál es la probabilidad de que la resistencia a ceder sea cuando mucho de 40? ¿De más de 60?

b. ¿Qué valor de resistencia a ceder separa al 75% más resistente del resto?

41. El dispositivo de apertura automática de un paracaídas de carga militar se diseñó para que lo abriera a 200 m sobre el suelo. Suponga que la altitud de abertura en realidad tiene una distribución normal con valor medio de 200 m y desviación estándar de 30 m. La carga útil se dañará si el paracaídas se abre a una altitud de menos de 100 m. ¿Cuál es la probabilidad de que se dañe la carga útil de cuando menos uno de cinco paracaídas lanzados en forma independiente?

42. La lectura de temperatura tomada con un termopar colocado en un medio a temperatura constante normalmente está distribuida con media μ, la temperatura real del medio, y desviación estándar σ. ¿Qué valor tendría σ para asegurarse de que el 95% de todas las lecturas están dentro de .1º de μ?

43. Se sabe que la distribución de resistencia de resistores de un tipo es normal y la resistencia del 10% de ellos es mayor de 10.256 ohms y la del 5% es de una resistencia menor de 9.671 ohms. ¿Cuáles son el valor medio y la desviación estándar de la distribución de resistencia?

44. Si la longitud roscada de un perno está normalmente distribuida, ¿cuál es la probabilidad de que la longitud roscada de un perno seleccionado al azar esté

a. dentro de 1.5 desviaciones estándar de su valor medio?

b. a más de 2.5 desviaciones estándar de su valor medio?

c. entre 1 y 2 desviaciones estándar de su valor medio?

45. Una máquina que produce cojinetes de bolas inicialmente se ajustó de modo que el diámetro promedio verdadero de los cojinetes que produce sea de .500 pulg. Un cojinete es aceptable si su diámetro está dentro de .004 pulg de su valor objetivo. Suponga, sin embargo, que el ajuste cambia durante el curso de la producción, de modo que los cojinetes tengan diámetros normalmente distribuidos con valor medio de .499 pulg y desviación estándar de .002 pulg. ¿Qué porcentaje de los cojinetes producidos no será aceptable?

46. La dureza Rockwell de un metal se determina hincando una punta endurecida en la superficie del metal y luego midiendo la profundidad de penetración de la punta. Suponga que la dureza Rockwell de una aleación particular está normalmente distribuida con media de 70 y desviación estándar de 3. (La dureza Rockwell se mide en una escala continua.)

a. Si una probeta es aceptable sólo si su dureza oscila entre 67 y 75, ¿cuál es la probabilidad de que una probeta seleccionada al azar tenga una dureza aceptable?

b. Si el rango de dureza aceptable es $(70 - c, 70 + c)$, ¿con qué valor de c tendría 95% de todas las probetas una dureza aceptable?

c. Si el rango de dureza aceptable es como el del inciso (a) y la dureza de cada una de diez probetas seleccionadas al azar se determina de forma independiente, ¿cuál es el valor esperado de probetas aceptables entre las diez?

d. ¿Cuál es la probabilidad de que cuando mucho ocho de diez probetas independientemente seleccionadas tengan una dureza de menos de 73.84? [*Sugerencia*: $Y =$ el número de entre las diez probetas con dureza de menos de 73.84 es una variable binomial; ¿cuál es p?]

47. La distribución de peso de paquetes enviados de cierta manera es normal con valor medio de 12 lb y desviación estándar de 3.5

lb. El servicio de paquetería desea establecer un valor de peso c más allá del cual habrá un cargo extra. ¿Qué valor de c es tal que 99% de todos los paquetes estén por lo menos 1 lb por debajo del peso de cargo extra?

48. Suponga que la tabla A.3 del apéndice contiene $\Phi(z)$ sólo para $z \geq 0$. Explique cómo aun así podría calcular
a. $P(-1.72 \leq Z \leq -.55)$
b. $P(-1.72 \leq Z \leq .55)$

¿Es necesario tabular $\Phi(z)$ para z negativo? ¿Qué propiedad de la curva normal estándar justifica su respuesta?

49. Considere los bebés nacidos en el rango "normal" de 37–43 semanas de gestación. Datos extensos sustentan la suposición de que el peso al nacer de estos bebés nacidos en Estados Unidos está normalmente distribuido con media de 3432 g y desviación estándar de 482 g. [El artículo "Are Babies Normal?" (*The American Statistician* (1999): 298–302) analizó datos de un año particular; para una selección sensible de intervalos de clase, un histograma no parecía del todo normal pero después de una investigación se determinó que esto se debía a que en algunos hospitales medían el peso en gramos, en otros lo medían a la onza más cercana y luego lo convertían en gramos. Una selección modificada de intervalos de clase que permitía esto produjo un histograma que era descrito muy bien por una distribución normal.]
a. ¿Cuál es la probabilidad de que el peso al nacer de un bebé seleccionado al azar de este tipo exceda de 4000 gramos? ¿Esté entre 3000 y 4000 gramos?
b. ¿Cuál es la probabilidad de que el peso al nacer de un bebé seleccionado al azar de este tipo sea de menos de 2000 gramos o de más de 5000 gramos?
c. ¿Cuál es la probabilidad de que el peso al nacer de un bebé seleccionado al azar de este tipo exceda de 7 libras?
d. ¿Cómo caracterizaría el .1% más extremo de todos los pesos al nacer?
e. Si X es una variable aleatoria con una distribución normal y a es una constante numérica ($a \neq 0$), entonces $Y = aX$ también tiene una distribución normal. Use esto para determinar la distribución de pesos al nacer expresados en libras (forma, media y desviación estándar) y luego calcule otra vez la probabilidad del inciso (c). ¿Cómo se compara ésta con su respuesta previa?

50. En respuesta a preocupaciones sobre el contenido nutricional de las comidas rápidas, McDonald's ha anunciado que utilizará un nuevo aceite de cocinar para sus papas a la francesa que reducirá sustancialmente los niveles de ácidos grasos e incrementará la cantidad de grasa poliinsaturada más benéfica. La compañía afirma que 97 de cada 100 personas no son capaces de detectar una diferencia de sabor entre los nuevos y los viejos aceites. Suponiendo que esta cifra es correcta (como proporción de largo plazo), ¿cuál es la probabilidad aproximada de que en una muestra aleatoria de 1000 individuos que han comprado papas a la francesa en McDonald's,
a. ¿Por lo menos 40 puedan notar la diferencia de sabor entre los dos aceites?
b. Cuando mucho 5% pueda notar la diferencia de sabor entre los dos aceites?

51. La desigualdad de Chebyshev (véase el ejercicio 44 del capítulo 3), es válida para distribuciones continuas y discretas. Estipula que para cualquier número k que satisfaga $k \geq 1$, $P(|X - \mu| \geq k\sigma) \leq 1/k^2$ (véase el ejercicio 44 en el capítulo 3 para una interpretación). Obtenga esta probabilidad en el caso de una distribución normal con $k = 1, 2, 3$ y compare con el límite superior.

52. Sea X el número de defectos en un carrete de cinta magnética de 100 m (una variable de valor entero). Suponga que X tiene aproximadamente una distribución normal con $\mu = 25$ y $\sigma = 5$. Use la corrección de continuidad para calcular la probabilidad de que el número de defectos sea
a. entre 20 y 30, inclusive
b. cuando mucho 30. Menos de 30.

53. Si X tiene una distribución binomial con parámetros $n = 25$ y p, calcule cada una de las siguientes probabilidades mediante la aproximación normal (con la corrección de continuidad) en los casos $p = .5, .6$, y $.8$ y compare con las probabilidades exactas calculadas con la tabla A.1 del apéndice.
a. $P(15 \leq X \leq 20)$
b. $P(X \leq 15)$
c. $P(20 \leq X)$

54. Suponga que 10% de todas las flechas de acero producidas por medio de un proceso no cumplen con las especificaciones pero pueden ser retrabajadas (en lugar de ser desechadas). Considere una muestra aleatoria de 200 flechas y sea X el número entre éstas que no cumplen con las especificaciones y pueden ser retrabajadas. ¿Cuál es la probabilidad aproximada de que X sea
a. cuando mucho 30?
b. menos que 30?
c. entre 15 y 25 (inclusive)?

55. Suponga que sólo 75% de todos los conductores en un estado usan con regularidad el cinturón de seguridad. Se selecciona una muestra aleatoria de 500 conductores. ¿Cuál es la probabilidad de que
a. entre 360 y 400 (inclusive) de los conductores en la muestra usen con regularidad el cinturón de seguridad?
b. menos de 400 de aquellos en la muestra usen con regularidad el cinturón de seguridad?

56. Demuestre que la relación entre un percentil normal general y el percentil z correspondiente es como se estipuló en esta sección.

57. **a.** Demuestre que si X tiene una distribución normal con parámetros μ y σ, entonces $Y = aX + b$ (una función lineal de X) también tiene una distribución normal. ¿Cuáles son los parámetros de la distribución de Y [es decir, $E(Y)$ y $V(Y)$]? [*Sugerencia*: escriba la función de distribución acumulativa de Y, $P(Y \leq y)$, como una integral que implique la función de densidad de probabilidad de X y luego diferencie con respecto a y para obtener la función de densidad de probabilidad de Y.]

b. Si cuando se mide en °C, la temperatura está normalmente distribuida con media de 115 y desviación estándar de 2, ¿qué se puede decir sobre la distribución de temperatura medida en °F?

58. No existe una fórmula exacta para función de distribución acumulativa normal estándar $\Phi(z)$, aunque se han publicado varias buenas aproximaciones en artículos. La siguiente se tomó de "Approximations for Hand Calculators Using Small Integer Coefficients" (*Mathematics of Computation*, 1977: 214–222). Con $0 < z \leq 5.5$,

$$P(Z \geq z) = 1 - \Phi(z) \approx .5 \exp\left\{-\left[\frac{(83z + 351)z + 562}{703/z + 165}\right]\right\}$$

El error relativo de esta aproximación es de menos de .042%. Úsela para calcular aproximaciones a las siguientes probabilidades y compare siempre que sea posible con las probabilidades obtenidas con la tabla A.3 del apéndice.

a. $P(Z \geq 1)$ **b.** $P(Z < -3)$
c. $P(-4 < Z < 4)$ **d.** $P(Z > 5)$

4.4 Distribuciones exponencial y gamma

La curva de densidad correspondiente a cualquier distribución normal tiene forma de campana y por consiguiente es simétrica. Existen muchas situaciones prácticas en las cuales la variable de interés para un investigador podría tener una distribución asimétrica. Una familia de distribuciones que tiene esta propiedad es la familia gamma. Primero se considera un caso especial, la distribución exponencial, y luego se le generaliza más adelante en esta sección.

Distribución exponencial

La familia de distribuciones exponenciales proporciona modelos de probabilidad que son muy utilizados en disciplinas de ingeniería y ciencia.

DEFINICIÓN

Se dice que X tiene una **distribución exponencial** con parámetro λ ($\lambda > 0$) si la función de densidad de probabilidad de X es

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{de lo contrario} \end{cases} \qquad (4.5)$$

Algunas fuentes escriben la función de densidad de probabilidad exponencial en la forma $(1/\beta)e^{-x/\beta}$, de modo que $\beta = 1/\lambda$. El valor esperado de una variable aleatoria X exponencialmente distribuida es

$$E(X) = \int_0^\infty x\lambda e^{-\lambda x}\, dx$$

Para obtener este valor esperado se requiere integrar por partes. La varianza de X se calcula utilizando el hecho de que $V(X) = E(X^2) - [E(X)]^2$. La determinación de $E(X^2)$ requiere integrar por partes dos veces en sucesión. Los resultados de estas integraciones son los siguientes:

$$\mu = \frac{1}{\lambda} \qquad \sigma^2 = \frac{1}{\lambda^2}$$

Tanto la media como la desviación estándar de la distribución exponencial son iguales a $1/\lambda$. En la figura 4.26 aparecen varias gráficas de varias funciones de densidad de probabilidad exponenciales.

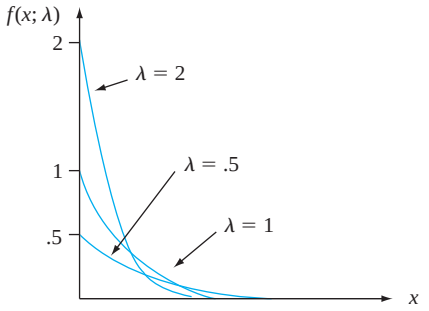


Figura 4.26 Curvas de densidad exponencial

La función de densidad de probabilidad exponencial es fácil de integrar para obtener la función de densidad acumulativa.

$$F(x;\lambda) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

Ejemplo 4.21 El artículo "Probabilistic Fatigue Evaluation of Riveted Railway Bridges" (*J. of Bridge Engr.*, 2008: 237–244) sugirió la distribución exponencial con valor medio de 6 MPa como modelo para la distribución del rango de esfuerzos en las conexiones de determinados puentes. Supongamos que éste es en realidad el verdadero modelo. Entonces $E(X) = 1/\lambda = 6$ implica que $\lambda = .1667$. La probabilidad de que el rango de esfuerzos a lo más sea de 10 MPa es

$$P(X \leq 10) = F(10; .1667) = 1 - e^{-(.1667)(10)} = 1 - .189 = .811$$

La probabilidad de que el rango de esfuerzo sea de entre 5 y 10 MPa es

$$P(5 \leq X \leq 10) = F(10; .1667) - F(5; .1667) = (1 - e^{-1.667}) - (1 - e^{-.8335}) = .246$$

■

La distribución exponencial se utiliza con frecuencia como modelo de la distribución de tiempos entre la ocurrencia de eventos sucesivos, tales como clientes que llegan a una instalación de servicio o llamadas que entran a un conmutador. La razón de esto es que la distribución exponencial está estrechamente relacionada con el proceso de Poisson discutido en el capítulo 3.

PROPOSICIÓN

Suponga que el número de eventos que ocurren en cualquier intervalo de tiempo de duración t tiene una distribución de Poisson con parámetro αt (donde α, la tasa del proceso de eventos, es el número esperado de eventos que ocurren en 1 unidad de tiempo) y que los números de ocurrencias en intervalos no traslapantes son independientes uno de otro. Entonces la distribución del tiempo transcurrido entre la ocurrencia de dos eventos sucesivos es exponencial con parámetro $\lambda = \alpha$.

Aunque una comprobación completa queda fuera del alcance de este libro, el resultado es fácil de verificar para el tiempo X_1 hasta que ocurre el primer evento:

$$P(X_1 \leq t) = 1 - P(X_1 > t) = 1 - P[\text{ningún evento en } (0, t)]$$
$$= 1 - \frac{e^{-\alpha t} \cdot (\alpha t)^0}{0!} = 1 - e^{-\alpha t}$$

la cual es exactamente la función de distribución acumulativa de la distribución exponencial.

Ejemplo 4.22 Suponga que se reciben llamadas durante las 24 horas en una "línea de emergencia para prevención del suicidio" de acuerdo con un proceso de Poisson a razón de $\alpha = .5$ llamadas por día. Entonces el número de días X entre llamadas sucesivas tiene una distribución exponencial con valor de parámetro .5, así que la probabilidad de que transcurran más de dos días entre llamadas es

$$P(X > 2) = 1 - P(X \le 2) = 1 - F(2; .5) = e^{-(.5)(2)} = .368$$

El tiempo esperado entre llamadas sucesivas es $1/.5 = 2$ días. ■

Otra aplicación importante de la distribución exponencial es modelar la distribución de la duración de un componente. Una razón parcial de la popularidad de tales aplicaciones es la **propiedad "de no memoria"** de la distribución exponencial. Suponga que la duración de un componente está exponencialmente distribuida con parámetro λ. Después de poner el componente en servicio, se deja que pase un periodo de t_0 horas y luego se ve si el componente sigue trabajando; ¿cuál es ahora la probabilidad de que dure por lo menos t horas más? En símbolos, se desea $P(X \ge t + t_0 | X \ge t_0)$. Por la definición de probabilidad condicional,

$$P(X \ge t + t_0 | X \ge t_0) = \frac{P[(X \ge t + t_0) \cap (X \ge t_0)]}{P(X \ge t_0)}$$

Pero el evento $X \ge t_0$ en el numerador es redundante, puesto que ambos eventos pueden ocurrir si y sólo si $X \ge t + t_0$. Por consiguiente,

$$P(X \ge t + t_0 | X \ge t_0) = \frac{P(X \ge t + t_0)}{P(X \ge t_0)} = \frac{1 - F(t + t_0; \lambda)}{1 - F(t_0; \lambda)} = e^{-\lambda t}$$

Esta probabilidad condicional es idéntica a la probabilidad original $P(X \ge t)$ de que el componente dure t horas. Por lo tanto *la distribución de duración adicional es exactamente la misma que la distribución original de duración*, así que en cada punto en el tiempo el componente no muestra ningún efecto de desgaste. En otras palabras, la distribución de la duración restante es independiente de la antigüedad actual.

Aunque la propiedad de no memoria se justifica por lo menos en forma aproximada en muchos problemas de aplicación, en otras situaciones los componentes se deterioran con el tiempo o de vez en cuando mejoran con él (por lo menos hasta cierto punto). Las distribuciones gamma, de Weibull y lognorma proporcionan modelos de duración más generales (las últimas dos se discuten en la siguiente sección).

La función gamma

Para definir la familia de distribuciones gamma, primero se tiene que introducir una función que desempeña un importante papel en muchas ramas de las matemáticas.

DEFINICIÓN

Con $\alpha > 0$, la **función gamma** $\Gamma(\alpha)$ se define como

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx \qquad (4.6)$$

Las propiedades más importantes de la función gamma son las siguientes:

1. Con cualquier $\alpha > 1$, $\Gamma(\alpha) = (\alpha - 1) \cdot \Gamma(\alpha - 1)$ [vía integración por partes]
2. Con cualquier entero positivo, n, $\Gamma(n) = (n - 1)!$
3. $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

De acuerdo con la expresión (4.6), si

$$f(x;\alpha) = \begin{cases} \dfrac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)} & x \geq 0 \\ 0 & \text{de lo contrario} \end{cases} \tag{4.7}$$

entonces $f(x;\alpha) \geq 0$ y $\int_0^\infty f(x;\alpha)\,dx = \Gamma(\alpha)/\Gamma(\alpha) = 1$, así que $f(x;\alpha)$ satisface las dos propiedades básicas de una función de densidad de probabilidad.

La distribución gamma

DEFINICIÓN

Se dice que una variable aleatoria continua X tiene una **distribución gamma** si la función de densidad de probabilidad de X es

$$f(x;\alpha,\beta) = \begin{cases} \dfrac{1}{\beta^\alpha\Gamma(\alpha)}x^{\alpha-1}e^{-x/\beta} & x \geq 0 \\ 0 & \text{de lo contrario} \end{cases} \tag{4.8}$$

donde los parámetros α y β satisfacen $\alpha > 0$, $\beta > 0$. La **distribución gamma estándar** tiene $\beta = 1$, así que (4.7) da la función de densidad de probabilidad de una variable aleatoria gamma estándar.

La distribución exponencial es resultado de considerar $\alpha = 1$ y $\beta = 1/\lambda$.

La figura 4.27(a) ilustra las gráficas de la función de densidad de probabilidad gamma $f(x;\alpha,\beta)$ (4.8) para varios pares (α,β), en tanto que la figura 4.27(b) presenta gráficas de la función de densidad de probabilidad gamma estándar. Para la función de densidad de probabilidad estándar cuando $\alpha \leq 1$, $f(x;\alpha)$ es estrictamente decreciente a medida que x se incrementa desde 0; cuando $\alpha > 1$, $f(x;\alpha)$ se eleva desde 0 en $x = 0$ hasta un máximo y luego decrece. El parámetro β en (4.8) se llama *parámetro de escala* porque los valores diferentes de 1 alargan o comprimen la función de densidad de probabilidad en la dirección x.

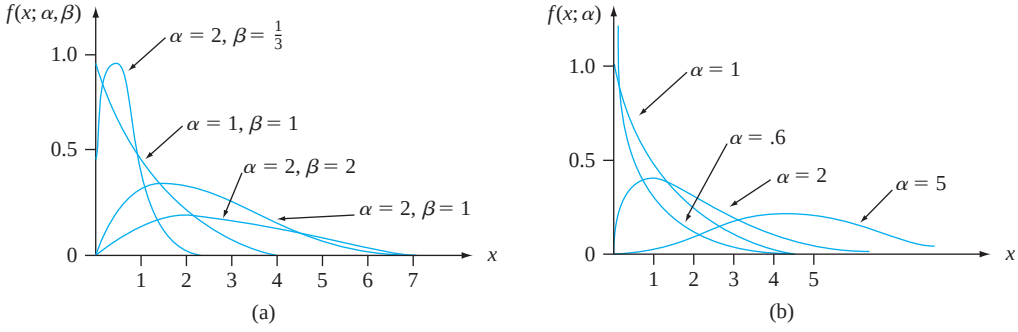


Figura 4.27 (a) Curvas de densidad gamma; (b) Curvas de densidad gamma estándar

La media y la varianza de una variable aleatoria X que tiene la distribución gamma $f(x;\alpha,\beta)$ son

$$E(X) = \mu = \alpha\beta \qquad V(X) = \sigma^2 = \alpha\beta^2$$

Cuando X es una variable aleatoria gamma estándar, la función de distribución acumulativa de X,

$$F(x;\alpha) = \int_0^x \frac{y^{\alpha-1}e^{-y}}{\Gamma(\alpha)}\,dy \qquad x > 0 \tag{4.9}$$

se llama **función gamma incompleta** [en ocasiones la función gamma incompleta se refiere a la expresión (4.9) sin el denominador $\Gamma(\alpha)$ en el integrando]. Existen tablas extensas de $F(x; \alpha)$ disponibles; en la tabla A.4 del apéndice se presenta una pequeña tabulación para $\alpha = 1, 2, \ldots, 10$ y $x = 1, 2, \ldots, 15$.

Ejemplo 4.23 Suponga que el tiempo de reacción X de un individuo seleccionado al azar a un estímulo tiene una distribución gamma estándar con $\alpha = 2$. Como

$$P(a \le X \le b) = F(b) - F(a)$$

cuando X es continua,

$$P(3 \le X \le 5) = F(5; 2) - F(3; 2) = .960 - .801 = .159$$

La probabilidad de que el tiempo de reacción sea de más de 4 s es

$$P(X > 4) = 1 - P(X \le 4) = 1 - F(4; 2) = 1 - .908 = .092$$ ■

La función gamma incompleta también se utiliza para calcular probabilidades que implican distribuciones gamma no estándar. Estas probabilidades también se obtienen casi instantáneamente con varios paquetes de software.

PROPOSICIÓN

Si X tiene una distribución gamma con parámetros α y β, entonces con cualquier $x > 0$, la función de distribución acumulativa de X es

$$P(X \le x) = F(x; \alpha, \beta) = F\left(\frac{x}{\beta}; \alpha\right)$$

donde $F(\cdot\,; \alpha)$ es la función gamma incompleta.

Ejemplo 4.24 Suponga que el tiempo de sobrevivencia X en semanas de un ratón macho seleccionado al azar expuesto a 240 rads de radiación gamma tiene una distribución gamma con $\alpha = 8$ y $\beta = 15$. (Datos en *Survival Distributions*: *Reliability Applications in the Biomedical Services*, de A. J. Gross y V. Clark, sugiere $\alpha \approx 8.5$ y $\beta \approx 13.3$.) El tiempo de sobrevivencia esperado es $E(X) = (8)(15) = 120$ semanas, en tanto que $V(X) = (8)(15)^2 = 1800$ y $\sigma_X = \sqrt{1800} = 42.43$ semanas. La probabilidad de que un ratón sobreviva entre 60 y 120 semanas es

$$\begin{aligned} P(60 \le X \le 120) &= P(X \le 120) - P(X \le 60) \\ &= F(120/15; 8) - F(60/15; 8) \\ &= F(8;8) - F(4;8) = .547 - .051 = .496 \end{aligned}$$

La probabilidad de que un ratón sobreviva por lo menos 30 semanas es

$$\begin{aligned} P(X \ge 30) &= 1 - P(X < 30) = 1 - P(X \le 30) \\ &= 1 - F(30/15; 8) = .999 \end{aligned}$$ ■

Distribución ji cuadrada

La distribución ji cuadrada es importante porque es la base de varios procedimientos de inferencia estadística. El papel central desempeñado por la distribución ji cuadrada en inferencia se deriva de su relación con distribuciones normales (véase el ejercicio 71). Se discutirá esta distribución con más detalle en capítulos posteriores.

DEFINICIÓN

Sea ν un entero positivo. Se dice entonces que una variable aleatoria X tiene una **distribución chi cuadrada** con parámetro ν si la función de densidad de probabilidad de X es la densidad gamma con $\alpha = \nu/2$ y $\beta = 2$. La función de densidad de probabilidad de una variable aleatoria ji cuadrada es por lo tanto

$$f(x; \nu) = \begin{cases} \dfrac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{(\nu/2)-1} e^{-x/2} & x \geq 0 \\ 0 & x < 0 \end{cases} \qquad (4.10)$$

El parámetro ν se llama **número de grados de libertad** (gl) de X. A menudo se utiliza el símbolo χ^2 en lugar de "ji cuadrada".

EJERCICIOS Sección 4.4 (59–71)

59. Sea X = el tiempo entre dos llegadas sucesivas a la ventanilla de autopago de un banco local. Si X tiene una distribución exponencial con $\lambda = 1$ (la cual es idéntica a una distribución gamma estándar con $\alpha = 1$), calcule lo siguiente:
- **a.** El tiempo esperado entre dos llegadas sucesivas.
- **b.** La desviación estándar del tiempo entre llegadas sucesivas
- **c.** $P(X \leq 4)$ **d.** $P(2 \leq X \leq 5)$

60. Sea X la distancia (m) que un animal recorre desde el sitio de su nacimiento hasta el primer territorio vacante que encuentra. Suponga que para ratas canguro con etiqueta en la cola, X tiene una distribución exponencial con parámetro $\lambda = .01386$ (como lo sugiere el artículo "Competition and Dispersal from Multiple Nests", *Ecology*, 1997: 873–883).
- **a.** ¿Cuál es la probabilidad de que la distancia sea cuando mucho de 100 m? ¿Cuando mucho de 200 m? ¿Entre 100 y 200 m?
- **b.** ¿Cuál es la probabilidad de que la distancia exceda la distancia media por más de 2 desviaciones estándar?
- **c.** ¿Cuál es el valor de la distancia mediana?

61. Los datos recogidos en el Aeropuerto Internacional Toronto Pearson sugiere que una distribución exponencial con valor medio de 2.725 horas es un buen modelo para la duración de la lluvia (*Urban Stormwater Management Planning with Analytical Probabilistic Models*, 2000, p. 69).
- **a.** ¿Cuál es la probabilidad de que la duración de un evento de lluvia en este lugar particular, sea por lo menos 2 horas? ¿A lo más 3 horas? ¿Entre 2 y 3 horas?
- **b.** ¿Cuál es la probabilidad de que la duración de la lluvia supere el valor medio por más de dos desviaciones estándar? ¿Cuál es la probabilidad de que sea menor que el valor medio en más de una desviación estándar?

62. El artículo "Microwave Observations of Daily Antarctic Sea-Ice Edge Expansion and Contribution Rates" *(IEEE Geosci. and Remote Sensing Letters*, 2006: 54–58) establece que "la distribución del avance-retroceso diarios del hielo marino con respecto a cada sensor es similar y es aproximadamente una exponencial doble". La distribución exponencial doble propuesta tiene una función de densidad $f(x) = .5\lambda e^{-\lambda|x|}$ con $-\infty < x < \infty$. La desviación estándar se da como 40.9 km.
- **a.** ¿Cuál es el valor del parámetro λ?
- **b.** ¿Cuál es la probabilidad de que la extensión del cambio del hielo marino esté dentro de 1 desviación estándar del valor medio?

63. Un consumidor está tratando de decidir entre dos planes de llamadas de larga distancia. El primero aplica una sola tarifa de 10¢ por minuto, en tanto que el segundo cobra una tarifa de 99¢ por llamadas hasta de 20 minutos y luego 10¢ por cada minuto adicional que exceda de 20 (suponga que las llamadas que duran un número no entero de minutos son cobradas proporcionalmente a un cargo por minuto entero). Suponga que la distribución de duración de llamadas del consumidor es exponencial con parámetro λ.
- **a.** Explique intuitivamente cómo la selección del plan de llamadas deberá depender de cuál sea la duración de las llamadas.
- **b.** ¿Cuál plan es mejor si la duración esperada de las llamadas es de 10 minutos? ¿Y de 15 minutos? [*Sugerencia*: sea $h_1(x)$ el costo del primer plan cuando la duración de las llamadas es de x minutos y sea $h_2(x)$ la función de costo del segundo plan. Dé expresiones para estas dos funciones de costo y luego determine el costo esperado de cada plan.]

64. Evalúe lo siguiente:
- **a.** $\Gamma(6)$ **b.** $\Gamma(5/2)$
- **c.** $F(4; 5)$ (la función gamma incompleta)
- **d.** $F(5; 4)$ **e.** $F(0; 4)$

65. Si X tiene una distribución gamma estándar con $\alpha = 7$, evalúe lo siguiente:
- **a.** $P(X \leq 5)$ **b.** $P(X < 5)$ **c.** $P(X > 8)$
- **d.** $P(3 \leq X \leq 8)$ **e.** $P(3 < X < 8)$
- **f.** $P(X < 4 \text{ o } X > 6)$

66. Suponga que el tiempo empleado por un estudiante seleccionado al azar que utiliza una terminal conectada a un sistema de computadoras de tiempo compartido tiene una distribución gamma con media de 20 min y varianza de 80 min^2.
- **a.** ¿Cuáles son los valores de α y β?
- **b.** ¿Cuál es la probabilidad de que un estudiante utilice la terminal durante cuando mucho 24 min?
- **c.** ¿Cuál es la probabilidad de que un estudiante utilice la terminal durante entre 20 y 40 min?

67. Suponga que cuando un transistor de cierto tipo se somete a una prueba de duración acelerada, la duración X (en semanas) tiene una distribución gamma con media de 24 semanas y desviación estándar de 12 semanas.

a. ¿Cuál es la probabilidad de que un transistor dure entre 12 y 24 semanas?

b. ¿Cuál es la probabilidad de que un transistor dure cuando mucho 24 semanas? ¿Es la mediana de la distribución de duración menor que 24? ¿Por qué sí o por qué no?

c. ¿Cuál es el 99º percentil de la distribución de duración?

d. Suponga que la prueba termina en realidad después de t semanas. ¿Qué valor de t es tal que sólo el .5% de todos los transistores continuarán funcionando al término?

68. El caso especial de la distribución gamma en la cual α es un entero positivo n se llama distribución de Erlang. Si se reemplaza β por $1/\lambda$ en la expresión (4.8), la función de densidad de probabilidad de Erlang es

$$f(x;\lambda,n)=\begin{cases}\dfrac{\lambda(\lambda x)^{n-1}e^{-\lambda x}}{(n-1)!} & x\ge 0\\ 0 & x<0\end{cases}$$

Se puede demostrar que si los tiempos entre eventos sucesivos son independientes, cada uno con distribución exponencial con parámetro λ, entonces el tiempo total que transcurre antes de que ocurran los siguientes n eventos tiene una función de densidad de probabilidad $f(x;\lambda,n)$.

a. ¿Cuál es el valor esperado de X? Si el tiempo (en minutos) entre llegadas de clientes sucesivos está exponencialmente distribuido con $\lambda = .5$, ¿cuánto tiempo se puede esperar que transcurra antes de que llegue el décimo cliente?

b. Si el tiempo entre llegadas de clientes está exponencialmente distribuido con $\lambda = .5$, ¿cuál es la probabilidad de que el décimo cliente (después del que acaba de llegar) llegue dentro de los siguientes 30 min?

c. El evento $\{X \le t\}$ ocurre si al menos ocurren n eventos en las siguientes t unidades de tiempo. Use el hecho de que el número de eventos que ocurren en un intervalo de duración t tiene una distribución de Poisson con parámetro λt para escribir una expresión (que implique probabilidades de Poisson) para la función de distribución acumulativa $F(t;\lambda,n) = P(X \le t)$.

69. Un sistema consta de cinco componentes idénticos conectados en serie como se muestra:

En cuanto un componente falla, todo el sistema lo hace. Suponga que cada componente tiene una duración que está exponencialmente distribuida con $\lambda = .01$ y que los componentes fallan de manera independiente uno de otro. Defina los eventos $A_i = \{$el componente i-ésimo dura por lo menos t horas$\}$, $i = 1, \ldots, 5$, de modo que los A_i son eventos independientes. Sea $X =$ el tiempo en el cual el sistema falla; es decir, la duración más corta (mínima) entre los cinco componentes.

a. ¿A qué evento equivale el evento $\{X \ge t\}$ que implique $A_1, \ldots, A_5$?

b. Utilizando la independencia de los eventos A_i, calcule $P(X \ge t)$. Luego obtenga $F(t) = P(X \le t)$ y la función de densidad de probabilidad de X. ¿Qué tipo de distribución tiene X?

c. Suponga que existen n componentes y cada uno tiene una duración exponencial con parámetro λ. ¿Qué tipo de distribución tiene X?

70. Si X tiene una distribución exponencial con parámetro λ, deduzca una expresión general para el $(100p)$º percentil de la distribución. Luego especialícela para obtener la mediana.

71. **a.** ¿A qué evento equivale el evento $\{X^2 \le y\}$ que implique a la X misma?

b. Si X tiene una distribución normal estándar, use el inciso (a) para escribir la integral que es igual a $P(X^2 \le y)$. Luego diferénciela con respecto a y para obtener la función de densidad de probabilidad de X^2 [el cuadrado de una variable $N(0, 1)$]. Por último, demuestre que X^2 tiene una distribución ji cuadrada con $\nu = 1$ grado de libertad [véase (4.10)]. [*Sugerencia*: use la siguiente identidad.]

$$\frac{d}{dy}\left\{\int_{a(y)}^{b(y)} f(x)\,dx\right\} = f[b(y)]\cdot b'(y) - f[a(y)]\cdot a'(y)$$

4.5 Otras distribuciones continuas

Las familias de distribuciones normal, gamma (incluida la exponencial) y uniforme proporcionan una amplia variedad de modelos de probabilidad de variables continuas, pero existen muchas situaciones prácticas en las cuales ningún miembro de estas familias se adapta bien a un conjunto de datos observados. Los estadísticos y otros investigadores han desarrollado otras familias de distribuciones que a menudo son apropiadas en la práctica.

Distribución de Weibull

El físico sueco Waloddi Weibull introdujo la familia de distribuciones Weibull en 1939; su artículo de 1951 "A Statistical Distribution Function of Wide Applicability" (*J. of Applied Mechanics*, vol. 18: 293–297) discute varias aplicaciones.

DEFINICIÓN

Se dice que una variable aleatoria X tiene una **distribución de Weibull** con parámetros α y β ($\alpha > 0, \beta > 0$) si la función de densidad de probabilidad de X es

$$f(x;\alpha,\beta) = \begin{cases} \dfrac{\alpha}{\beta^{\alpha}} x^{\alpha-1} e^{-(x/\beta)^{\alpha}} & x \geq 0 \\ 0 & x < 0 \end{cases} \qquad (4.11)$$

En algunas situaciones, existen justificaciones teóricas para la pertinencia de la distribución de Weibull, pero en muchas aplicaciones $f(x; \alpha, \beta)$ simplemente proporciona una concordancia con los datos observados con valores particulares de α y β. Cuando $\alpha = 1$, la función de densidad de probabilidad se reduce a la distribución exponencial (con $\lambda = 1/\beta$), de modo que la distribución exponencial es un caso especial tanto de la distribución gamma como de la distribución de Weibull. No obstante, existen distribuciones gamma que no son Weibull, y viceversa, por lo que una familia no es un subconjunto de la otra. Tanto α como β pueden ser variadas para obtener diferentes formas de curvas de densidad, como se ilustra en la figura 4.28. β es un parámetro de escala, así que diferentes valores alargan o comprimen la gráfica en la dirección x y α es un parámetro de la forma de la curva.

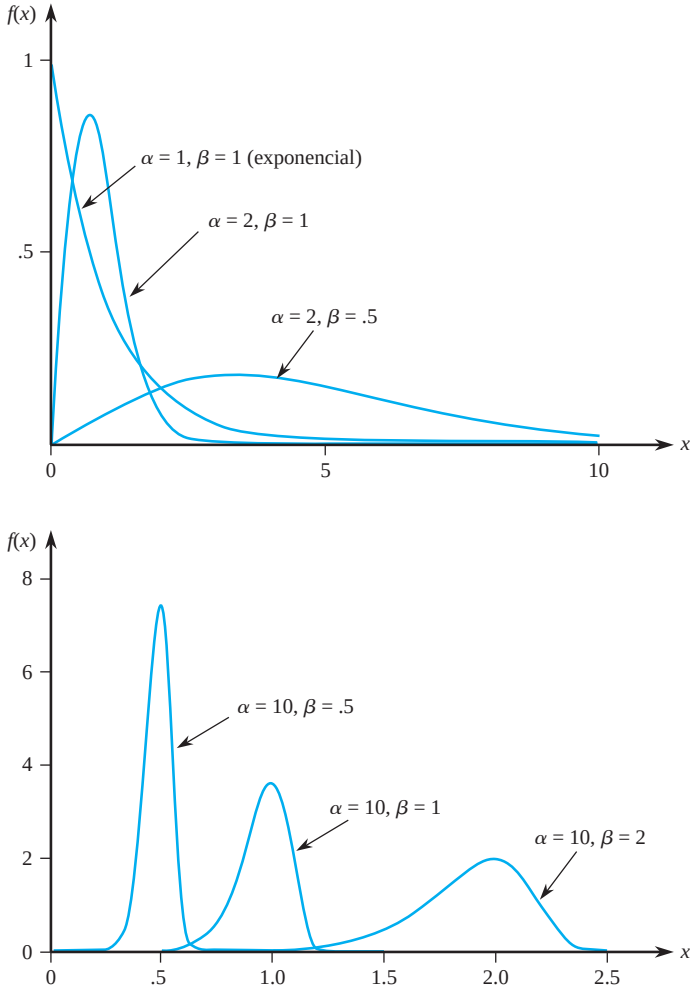


Figura 4.28 Curvas de densidad de Weibull

Si se integra para obtener $E(X)$ y $E(X^2)$ se tiene

$$\mu = \beta\Gamma\left(1 + \frac{1}{\alpha}\right) \qquad \sigma^2 = \beta^2\left\{\Gamma\left(1 + \frac{2}{\alpha}\right) - \left[\Gamma\left(1 + \frac{1}{\alpha}\right)\right]^2\right\}$$

El cálculo de μ y σ^2 requiere por lo tanto el uso de la función gamma.

La integración $\int_0^x f(y; \alpha, \beta)\, dy$ es fácil de realizar para obtener la función de distribución acumulativa de X.

La función de distribución acumulativa de una variable aleatoria de Weibull con parámetros α y β es

$$F(x; \alpha, \beta) = \begin{cases} 0 & x < 0 \\ 1 - e^{-(x/\beta)^\alpha} & x \geq 0 \end{cases} \qquad (4.12)$$

Ejemplo 4.25 En años recientes la distribución de Weibull ha sido utilizada para modelar emisiones de varios contaminantes por motores. Sea X la cantidad de emisiones de NO_x (g/gal) de un motor de cuatro tiempos de un tipo seleccionado al azar, y suponga que X tiene una distribución de Weibull con $\alpha = 2$ y $\beta = 10$ (sugeridos por la información que aparece en el artículo "Quantification of Variability and Uncertainty in Lawn and Garden Equipment NO_x and Total Hydrocarbon Emission Factors", *J. of the Air and Waste Management Assoc.*, 2002: 435–448). La curva de densidad correspondiente se ve exactamente como la de la figura 4.28 con $\alpha = 2$, $\beta = 1$, excepto que ahora los valores 50 y 100 reemplazan a 5 y 10 en el eje horizontal (debido a que β es un "parámetro de escala"). Entonces

$$P(X \leq 10) = F(10; 2, 10) = 1 - e^{-(10/10)^2} = 1 - e^{-1} = .632$$

Asimismo, $P(X \leq 25) = .998$, así que la distribución está concentrada casi por completo en valores entre 0 y 25. El valor c, el cual separa 5% de todos los motores que emiten las más grandes cantidades de NO_x del 95% restante, satisface

$$.95 = 1 - e^{-(c/10)^2}$$

Aislando el término exponencial en un lado, sacando logaritmos y resolviendo la ecuación resultante se obtiene $c \approx 17.3$ como el 95º percentil de la distribución de emisiones. ■

En situaciones prácticas, un modelo de Weibull puede ser razonable excepto que el valor de X más pequeño posible puede ser algún valor γ que no se supuso fuera cero (esto también se aplicaría a un modelo gamma). La cantidad γ puede entonces ser considerada como un tercer (umbral) parámetro de la distribución, lo cual es lo que Weibull hizo en su trabajo original. Con, por ejemplo, $\gamma = 3$, todas las curvas que aparecen en la figura 4.28 se desplazarían 3 unidades a la derecha. Esto equivale a decir que $X - \gamma$ tiene la función de densidad de probabilidad (4.11) de modo que la función de distribución acumulativa de X se obtiene reemplazando x en (4.12) por $x - \gamma$.

Ejemplo 4.26 La comprensión de las propiedades volumétricas del asfalto es importante en el diseño de mezclas que se traducirán en pavimento de alta durabilidad. El artículo "Is a Normal Distribution the Most Appropriate Statistical Distribution for Volumetric Properties in Asphalt Mixtures?" (*J. of Testing and Evaluation*, sept. 2009: 1–11) utilizó el análisis de algunos datos de la muestra para recomendar que para una mezcla particular, X = volumen vacío de aire (%) se modela con una distribución de Weibull de tres parámetros. Supongamos los valores de los parámetros $\gamma = 4$, $\alpha = 1.3$ y $\beta = .8$ (muy cerca de las estimaciones hechas en el artículo).

Para $x > 4$, la función de distribución acumulativa es

$$F(x; \alpha, \beta, \gamma) = F(x; 1.3, .8, 4) = 1 - e^{-[(x-4)/.8]^{1.3}}$$

La probabilidad de que el volumen vacío de aire de una muestra está entre 5% y 6% es

$$P(5 \le X \le 6) = F(6; 1.3,.8,4) - F(5; 1.3, .8, 4) = e^{-[(5-4)/.8]^{1.3}} - e^{-[(6-4)/.8]^{1.3}}$$
$$= .263 - .037 = .226$$

La figura 4.29 muestra una gráfica de Minitab de la correspondiente función de densidad de Weibull en la que el área sombreada corresponde a la probabilidad que se acaba de calcular.

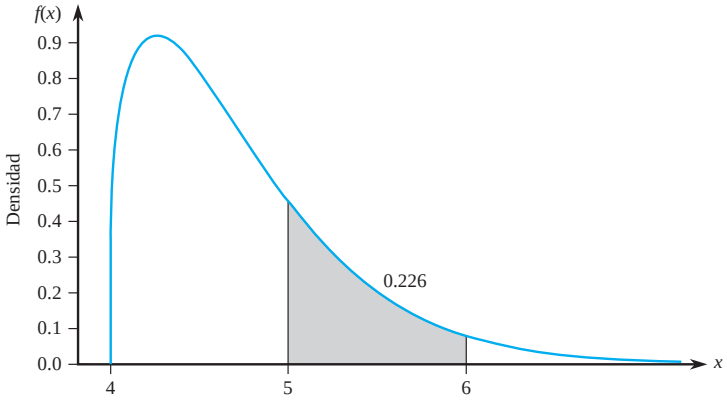


Figura 4.29 Curva de densidad de Weibull con umbral = 4, forma = 1.3, escala = .8 ■

Distribución lognormal

DEFINICIÓN

Se dice que una variable aleatoria no negativa X tiene una **distribución lognormal** si la variable aleatoria $Y = \ln(X)$ tiene una distribución normal. La función de densidad de probabilidad resultante de una variable aleatoria lognormal cuando el $\ln(X)$ está normalmente distribuido con parámetros μ y σ es

$$f(x; \mu, \sigma) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\sigma x} e^{-[\ln(x)-\mu]^2/(2\sigma^2)} & x \ge 0 \\ 0 & x < 0 \end{cases}$$

Hay que tener cuidado aquí; los parámetros μ y σ no son la media y la desviación estándar de X sino de $\ln(X)$. Es común referirse a μ y σ como los parámetros de ubicación y de escala, respectivamente. La media y la varianza de X se puede demostrar que son

$$E(X) = e^{\mu+\sigma^2/2} \qquad V(X) = e^{2\mu+\sigma^2} \cdot (e^{\sigma^2} - 1)$$

En el capítulo 5 se presenta una justificación teórica para esta distribución en conexión con el teorema del límite central, pero como con cualesquiera otras distribuciones, se puede utilizar la lognormal como modelo incluso en la ausencia de semejante justificación. La figura 4.30 ilustra gráficas de la función de densidad de probabilidad lognormal; aunque una curva normal es simétrica, una curva lognormal tiene una asimetría positiva.

Como el $\ln(X)$ tiene una distribución normal, la función de distribución acumulativa de X puede ser expresada en términos de la función de distribución acumulativa $\Phi(z)$ de una variable aleatoria normal estándar Z.

$$F(x; \mu, \sigma) = P(X \le x) = P[\ln(X) \le \ln(x)]$$
$$= P\left(Z \le \frac{\ln(x) - \mu}{\sigma}\right) = \Phi\left(\frac{\ln(x) - \mu}{\sigma}\right) \quad x \ge 0 \qquad (4.13)$$

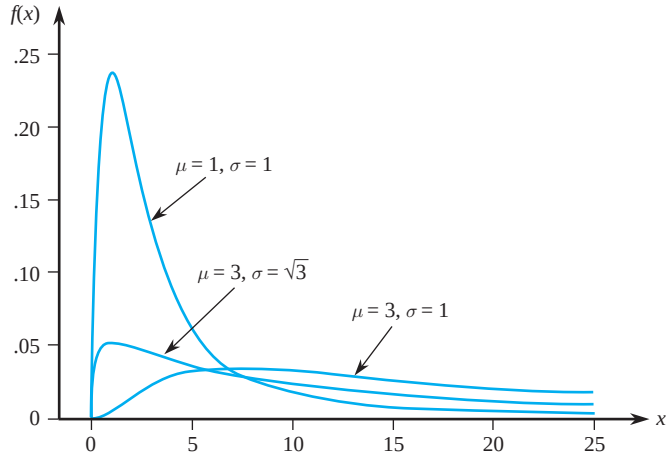


Figura 4.30 Curvas de densidad lognormal

Ejemplo 4.27 De acuerdo con el artículo "Predictive Model for Pitting Corrosion in Buried Oil and Gas Pipelines" (*Corrosion*, 2009: 332–342), la distribución logarítmica normal ha sido considerada como la mejor opción para describir la distribución de los datos de máxima profundidad de pozo de las tuberías de hierro fundido en el suelo. Los autores sugieren que una distribución logarítmica normal con $\mu = .353$ y $\sigma = .754$ es apropiada para la profundidad de pozo máxima (mm) de tuberías enterradas. Para esta distribución, el valor medio y la varianza de la profundidad del pozo son

$$E(X) = e^{.353+(.754)^2/2} = e^{.6373} = 1.891$$
$$V(X) = e^{2(.353)+(.754)^2} \cdot (e^{(.754)^2} - 1) = (3.57697)(.765645) = 2.7387$$

La probabilidad de que la máxima profundidad de pozo esté entre 1 y 2 mm es

$$P(1 \le X \le 2) = P(\ln(1) \le \ln(X) \le \ln(2)) = P(0 \le \ln(X) \le .693)$$
$$= P\left(\frac{0 - .353}{.754} \le Z \le \frac{.693 - .353}{.754}\right) = \Phi(.47) - \Phi(-.45) = .354$$

Esta probabilidad se ilustra en la figura 4.31 (de Minitab).

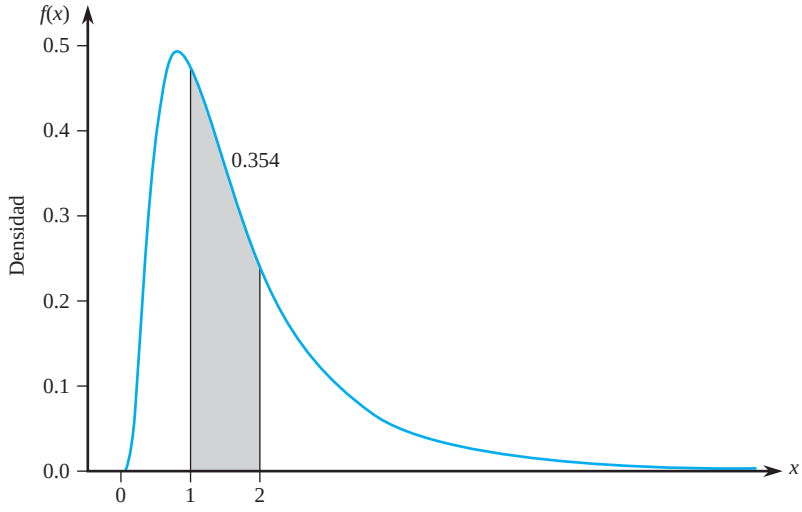


Figura 4.31 Curva de densidad lognormal con ubicación = .353 y escala = .754

¿Qué valor de c es tal que sólo el 1% de todas las muestras tienen una profundidad máxima de pozo que excede c? El valor que se desea es

$$.99 = P(X \le c) = P\left(Z \le \frac{\ln(c) - .353}{.754}\right)$$

El valor z crítico 2.33 captura un área de cola superior de .01 ($z_{.01} = 2.33$), y por tanto un área acumulada de .99. Esto implica que

$$\frac{\ln(c) - .353}{.754} = 2.33$$

con lo que $\ln(c) = 2.1098$ y $c = 8.247$. Por lo tanto, 8.247 es el 99º percentil de la distribución de la máxima profundidad de pozo. ■

Distribución beta

Todas las familias de distribuciones continuas estudiadas hasta ahora excepto la distribución uniforme tienen densidad positiva a lo largo de un intervalo infinito (aunque por lo general la función de densidad se reduce con rapidez a cero más allá de unas cuantas desviaciones estándar de la media). La distribución beta proporciona densidad positiva sólo para X en un intervalo de longitud finita.

DEFINICIÓN

Se dice que una variable aleatoria X tiene una **distribución beta** con parámetros α, β (ambos positivos), A y B si la función de densidad de probabilidad de X es

$$f(x; \alpha, \beta, A, B) = \begin{cases} \dfrac{1}{B-A} \cdot \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} \left(\dfrac{x-A}{B-A}\right)^{\alpha-1} \left(\dfrac{B-x}{B-A}\right)^{\beta-1} & A \le x \le B \\ 0 & \text{de lo contrario} \end{cases}$$

El caso $A = 0$, $B = 1$ da la **distribución beta estándar.**

La figura 4.32 ilustra varias funciones de densidad de probabilidad beta estándar. Las gráficas de la función de densidad de probabilidad son similares, excepto que están desplazadas y luego alargadas o comprimidas para ajustarse al intervalo $[A, B]$. A menos que α y β sean enteros, la integración de la función de densidad de probabilidad para calcular probabilidades es difícil. Se deberá utilizar una tabla de la función beta incompleta o un programa de computadora apropiado. La media y varianza de X son

$$\mu = A + (B - A) \cdot \frac{\alpha}{\alpha + \beta} \qquad \sigma^2 = \frac{(B-A)^2 \alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

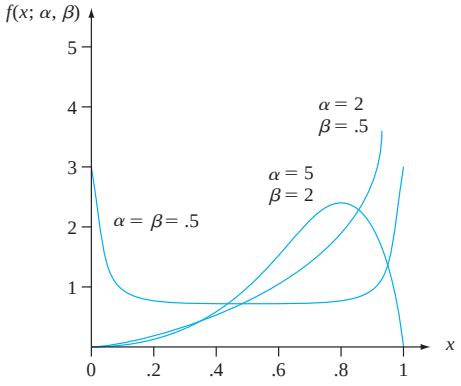


Figura 4.32 Curvas de densidad beta estándar

Ejemplo 4.28 Los gerentes de proyectos a menudo utilizan un método llamado PERT (por las siglas en inglés de técnica de revisión y evaluación de programas) para coordinar las diversas actividades que conforman un gran proyecto. (Una aplicación exitosa ocurrió en la construcción de la nave espacial *Apolo*.) Una suposición estándar en el análisis PERT es que el tiempo necesario para completar cualquier actividad particular una vez que se ha iniciado tiene una distribución beta con A = el tiempo optimista (si todo sale bien) y B = tiempo pesimista (si todo sale mal). Suponga que al construir una casa unifamiliar, el tiempo X (en días) necesario para echar los cimientos tiene una distribución beta con $A = 2$, $B = 5$, $\alpha = 2$ y $\beta = 3$. Entonces, $\alpha/(\alpha + \beta) = .4$, así que $E(X) = 2 + (3)(.4) = 3.2$. Con estos valores de α y β, la función de densidad de probabilidad de X es una función polinomial simple. La probabilidad de que se requieran a lo más tres días para echar los cimientos es

$$P(X \le 3) = \int_2^3 \frac{1}{3} \cdot \frac{4!}{1!2!}\left(\frac{x-2}{3}\right)\left(\frac{5-x}{3}\right)^2 dx$$
$$= \frac{4}{27}\int_2^3 (x-2)(5-x)^2 dx = \frac{4}{27}\cdot\frac{11}{4} = \frac{11}{27} = .407$$ ■

La distribución beta estándar se utiliza comúnmente para modelar la variación en la proporción o porcentaje de una cantidad que ocurre en diferentes muestras, tal como la proporción de un día de 24 horas que un individuo está despierto o la proporción de cierto elemento en un compuesto químico.

EJERCICIOS Sección 4.5 (72–86)

72. La duración X (en cientos de horas) de un tipo de tubo de vacío tiene una distribución de Weibull con parámetros $\alpha = 2$ y $\beta = 3$. Calcule lo siguiente:
- **a.** $E(X)$ y $V(X)$
- **b.** $P(X \le 6)$
- **c.** $P(1.5 \le X \le 6)$

(Esta distribución de Weibull se sugiere como modelo del tiempo de servicio en "On the Assessment of Equipment Reliability: Trading Data Collection Costs for Precision", *J. of Engr. Manuf.*, 1991: 105–109.)

73. Los autores del artículo "A Probabilistic Insulation Life Model for Combined Thermal-Electrical Stresses" (*IEEE Trans. on Elect. Insulation*, 1985: 519–522) expresan que "la distribución de Weibull se utiliza mucho en problemas estadísticos relacionados con el desgaste de materiales sólidos aislantes sometidos a envejecimiento y esfuerzo". Proponen el uso de la distribución como modelo del tiempo (en horas) hasta la falla de especímenes aislantes sólidos sometidos a voltaje de CA. Los valores de los parámetros dependen del voltaje y temperatura; suponga $\alpha = 2.5$ y $\beta = 200$ (valores sugeridos por datos que aparecen en el artículo).
- **a.** ¿Cuál es la probabilidad de que la duración de un espécimen sea cuando mucho de 250? ¿De menos de 250? ¿De más de 300?
- **b.** ¿Cuál es la probabilidad de que la duración de un espécimen sea de entre 100 y 250?
- **c.** ¿Qué valor es tal que exactamente 50% de todos los especímenes tengan duraciones que sobrepasen ese valor?

74. Sea X = el tiempo (en 10^{-1} semanas) desde el envío de un producto defectuoso hasta que el cliente lo devuelve. Suponga que el tiempo de devolución mínimo es $\gamma = 3.5$ y que el excedente $X - 3.5$ sobre el mínimo tiene una distribución de Weibull con parámetros $\alpha = 2$ y $\beta = 1.5$ (véase el artículo "Practical Applications of the Weibull Distribution", *Industrial Quality Control*, agosto de 1964: 71–78).
- **a.** ¿Cuál es la función de distribución acumulativa de X?
- **b.** ¿Cuáles son el tiempo de devolución esperado y la varianza del tiempo de devolución? [*Sugerencia*: primero obtenga $E(X - 3.5)$ y $V(X - 3.5)$.]
- **c.** Calcule $P(X > 5)$.
- **d.** Calcule $P(5 \le X \le 8)$.

75. Si X tiene una distribución de Weibull con la función de densidad de probabilidad de la expresión (4.11), verifique que $\mu = \beta\Gamma(1 + 1/\alpha)$. [*Sugerencia*: en la integral para $E(X)$ cambie la variable $y = (x/\beta)^\alpha$, de modo que $x = \beta y^{1/\alpha}$.]

76.
- **a.** En el ejercicio 72, ¿cuál es la duración mediana de los tubos? [*Sugerencia*: use la expresión (4.12).]
- **b.** En el ejercicio 74, ¿cuál es el tiempo de devolución mediano?
- **c.** Si X tiene una distribución de Weibull con la función de distribución acumulativa de la expresión (4.12), obtenga una expresión general para el percentil $(100p)^o$ de la distribución.
- **d.** En el ejercicio 74, la compañía desea negarse a aceptar devoluciones después de t semanas. ¿Para qué valor de t sólo el 10% de todas las devoluciones serán rechazadas?

77. Los autores del artículo del cual se extrajeron los datos en el ejercicio 1.27 sugirieron que un modelo de probabilidad razonable de la duración de las brocas era una distribución lognormal con $\mu = 4.5$ y $\sigma = .8$.
- **a.** ¿Cuáles son el valor medio y la desviación estándar de la duración?

b. ¿Cuál es la probabilidad de que la duración sea cuando mucho de 100?
c. ¿Cuál es la probabilidad de que la duración sea por lo menos de 200? ¿De más de 200?

78. El artículo "On Assessing the Accuracy of Offshore Wind Turbine Reliability-Based Design Loads from the Environmental Contour Method" (*Intl. J. of Offshore and Polar Engr.*, 2005: 132–140) propone la distribución de Weibull con $\alpha = 1.817$ y $\beta = .863$ como modelo de una altura (m) de olas significativa durante 1 hora en un sitio.
a. ¿Cuál es la probabilidad de que la altura de las olas sea cuando mucho de .5 m?
b. ¿Cuál es la probabilidad de que la altura de las olas exceda su valor medio por más de una desviación estándar?
c. ¿Cuál es la mediana de la distribución de la altura de las olas?
d. Para $0 < p < 1$, dé una expresión general para el percentil $100p$º de la distribución de altura de olas.

79. Cargas de fuentes no puntuales son masas de químicos que viajan al caudal principal de un río y sus afluentes, en flujos que se distribuyen sobre un flujo relativamente de largo alcance, a diferencia de los que entran en puntos bien definidos y regulados. El artículo "Assessing Uncertainty in Mass Balance Calculation of River Nonpoint Source Loads" (*J. of Envir. Engr.*, 2008: 247–258) sugirió que para cierto periodo y lugar, X = carga de fuentes no puntuales de sólidos disueltos totales podría ser modelada con una distribución logarítmica normal con valor medio de 10,281 kg/día/km y un coeficiente de variación $CV = .40$ ($CV = \sigma_X/\mu_X$).
a. ¿Cuáles es el valor medio y la desviación estándar de ln (X)?
b. ¿Cuál es la probabilidad de que X sea a lo sumo 15,000 kg/día/km?
c. ¿Cuál es la probabilidad de que X supere su valor medio y por qué esta probabilidad no es .5?
d. ¿Es 17,000 el percentil 95 de la distribución?

80. **a.** Use la ecuación (4.13) para escribir una fórmula para la mediana $\tilde{\mu}$ de la distribución lognormal. ¿Cuál es la mediana de la distribución de carga del ejercicio 79?
b. Recordando que z_α es la notación para el percentil $100(1 - \alpha)$ de la distribución normal estándar, escriba una expresión para el percentil $100(1 - \alpha)$ de la distribución lognormal. En el ejercicio 79, ¿qué valor excederá la carga recibida sólo 1% del tiempo?

81. Una justificación teórica basada en el mecanismo de falla de cierto material sustenta la suposición de que la resistencia dúctil X de un material tiene una distribución lognormal. Suponga que los parámetros son $\mu = 5$ y $\sigma = .1$.
a. Calcule $E(X)$ y $V(X)$.
b. Calcule $P(X > 125)$.
c. Calcule $P(110 \leq X \leq 125)$.
d. ¿Cuál es el valor de la resistencia dúctil mediana?
e. Si diez muestras diferentes de un acero de aleación de este tipo se sometieran a una prueba de resistencia, ¿cuántas esperaría que tengan una resistencia de por lo menos 125?
f. Si 5% de los valores de resistencia más pequeños fueran inaceptables, ¿cuál sería la resistencia mínima aceptable?

82. El artículo "The Statistics of Phytotoxic Air Pollutants" (*J. of Royal Stat. Soc.*, 1989:183–198) sugiere la distribución lognormal como modelo de la concentración de SO_2 sobre cierto bosque. Suponga que los valores de parámetro son $\mu = 1.9$ y $\sigma = .9$.
a. ¿Cuáles son el valor medio y la desviación estándar de la concentración?
b. ¿Cuál es la probabilidad de que la concentración sea cuando mucho de 10? ¿De entre 5 y 10?

83. ¿Qué condición en relación con α y β es necesaria para que la función de densidad de probabilidad beta estándar sea simétrica?

84. Suponga que la proporción X de área superficial en un cuadrado seleccionado al azar que está cubierto por cierta planta tiene una distribución beta estándar con $\alpha = 5$ y $\beta = 2$.
a. Calcule $E(X)$ y $V(X)$.
b. Calcule $P(X \leq .2)$.
c. Calcule $P(.2 \leq X \leq .4)$.
d. ¿Cuál es la proporción esperada de la región de muestreo no cubierta por la planta?

85. Sea X que tiene una densidad beta estándar con parámetros α y β.
a. Verifique la fórmula para $E(X)$ dada en la sección.
b. Calcule $E[(1 - X)^m]$. Si X representa la proporción de una sustancia compuesta de un ingrediente particular, ¿cuál es la proporción esperada que no se compone de ese ingrediente?

86. Se aplica esfuerzo a una barra de acero de 20 pulg sujeta por cada extremo en una posición fija. Sea Y = la distancia del extremo izquierdo al punto donde se rompe la barra. Suponga que $Y/20$ tiene una distribución beta estándar con $E(Y) = 10$ y $V(Y) = \frac{100}{7}$.
a. ¿Cuáles son los parámetros de la distribución beta estándar pertinente?
b. Calcule $P(8 \leq Y \leq 12)$.
c. Calcule la probabilidad de que la barra se rompa a más de 2 pulg de donde esperaba que se rompiera.

4.6 Gráficas de probabilidad

Un investigador a menudo ha obtenido una muestra numérica $x_1, x_2, \ldots, x_n$ y desea saber si es factible que provenga de una distribución de población de un tipo particular (p. ej., de una distribución normal). Entre otras cosas, muchos procedimientos formales de inferencia estadística están basados en la suposición de que la distribución de población es de un tipo específico. El uso de un procedimiento como ésos es inapropiado si la distribución de probabilidad subyacente real difiere en gran medida del tipo supuesto. Por ejemplo, el

artículo "Toothpaste Detergents: A Potential Source of Oral Soft Tissue Damage" (*Intl. J. of Dental Hygiene*, 2008: 193–198) contiene la siguiente declaración: "Debido a que el número de muestras para cada experimento (replicación) fue limitado a tres fuentes según el tipo de tratamiento, se supone que los datos están normalmente distribuidos". Como justificación de este acto de fe, los autores escribieron que "las estadísticas descriptivas mostraron desviaciones estándar que sugieren una distribución normal altamente probable". *Nota*: este argumento no es muy convincente.

Además, el entendimiento de la distribución subyacente en ocasiones puede dar una idea de los mecanismos físicos implicados en la generación de los datos. Una forma efectiva de verificar una suposición distribucional es construir una **gráfica de probabilidad**. La esencia de una gráfica como ésa es que si la distribución en la cual está basada es correcta, los puntos en la gráfica quedarán casi en una línea recta. Si la distribución real es bastante diferente de la utilizada para construir la gráfica, los puntos deberán apartarse sustancialmente de un patrón lineal.

Percentiles muestrales

Los detalles implicados al construir gráficas de probabilidad difieren un poco de una fuente a otra. La base de la construcción es una comparación entre percentiles de los datos muestrales y los percentiles correspondientes de la distribución considerada. Recuérdese que el percentil $(100p)^{\text{o}}$ de una distribución continua con función de distribución acumulativa $F(\cdot)$ es el número $\eta(p)$ que satisface $F(\eta(p)) = p$. Es decir, $\eta(p)$ es el número sobre la escala de medición de modo que el área bajo la curva de densidad a la izquierda de $\eta(p)$ es p. Por lo tanto, el percentil 50^{o}, $\eta(.5)$, satisface $F(\eta(.5)) = .5$, y el percentil 90^{o} satisface $F(\eta(.9)) = .9$. Considere como ejemplo la distribución normal estándar, para la cual la función de distribución acumulativa es $\Phi(\cdot)$. En la tabla A.3 del apéndice, el 20^{o} percentil se halla localizando la fila y columna en la cual aparece .2000 (o un número tan cerca de él como sea posible) en el interior de la tabla. Como .2005 aparece en la intersección de la fila $-.8$ y la columna .04, el 20^{o} percentil es aproximadamente $-.84$. Asimismo, el 25^{o} percentil de la distribución normal estándar es (utilizando interpolación lineal) aproximadamente $-.675$.

En general, los percentiles muestrales se definen del mismo modo que los percentiles de una distribución de población. El 50^{o} percentil muestral deberá separar el 50% más pequeño de la muestra del 50% más grande, el 90^{o} percentil deberá ser tal que el 90% de la muestra quede debajo de ese valor y el 10% quede sobre ese valor, y así de manera sucesiva. Desafortunadamente, se presentan problemas cuando en realidad se trata de calcular los percentiles muestrales de una muestra particular de n observaciones. Si, por ejemplo, $n = 10$, se puede separar 20% de estos valores o 30% de los datos, pero no hay ningún valor que separe con exactitud 23% de estas diez observaciones. Para ir más allá, se requiere una definición operacional de percentiles muestrales (éste es un lugar donde diferentes personas hacen cosas un poco diferentes). Recuérdese que cuando n es impar, la mediana muestral o el 50^{o} percentil muestral es el valor medio en la lista ordenada, por ejemplo, el sexto valor más grande cuando $n = 11$. Esto equivale a considerar la observación media como la mitad en la mitad inferior de los datos y la mitad en la mitad superior. Asimismo, supóngase $n = 10$. Entonces, si a este tercer valor más pequeño se le da el nombre de 25^{o} percentil, ese valor se está considerando como la mitad en el grupo inferior (compuesto de las dos observaciones más pequeñas) y la mitad en el grupo superior (las siete observaciones más grandes). Esto conduce a la siguiente definición general de percentiles muestrales.

DEFINICIÓN

Se ordenan las n observaciones muestrales de la más pequeña a la más grande. Entonces la observación i-ésima más pequeña en la lista se considera que es el **$[100(i - .5)/n]^{\text{o}}$ percentil muestral**.

Una vez que se han calculado los valores porcentuales $100(i - .5)/n$ ($i = 1, 2, \ldots, n$) se pueden obtener los percentiles muestrales correspondientes a porcentajes intermedios mediante interpolación lineal. Por ejemplo, si $n = 10$, los porcentajes correspondientes a las observaciones muestrales ordenadas son $100(1 - .5)/10 = 5\%$, $100(2 - .5)/10 = 15\%$, $25\% \ldots$, y $100(10 - .5)/10 = 95\%$. El 10º percentil está entonces a la mitad entre el 5º percentil (observación muestral más pequeña) y el 15º (segunda observación más pequeña). Para los propósitos de este libro, tal interpolación no es necesaria porque una gráfica de probabilidad se basa sólo en los porcentajes $100(i - .5)/n$ correspondientes a las n observaciones muestrales.

Gráfica de probabilidad

Supóngase ahora que para los porcentajes $100(i - .5)/n$ ($i = 1, \ldots, n$) se determinan los percentiles de una distribución de población especificada cuya factibilidad está siendo investigada. Si la muestra en realidad se seleccionó de la distribución especificada, los percentiles muestrales (observaciones muestrales ordenadas) deberán estar razonablemente próximos a los percentiles de distribución de población correspondientes. Es decir, con $i = 1, 2, \ldots, n$ deberá haber una razonable concordancia entre la i-ésima observación muestral más pequeña y el $[100(i - .5)/n]$º percentil de la distribución especificada. Considérense los pares (percentil poblacional, percentil muestral); es decir, los pares

$$\left(\begin{matrix}[100(i - .5)/n]\text{° percentil} \\ \text{de la distribución,}\end{matrix}, \begin{matrix}i\text{-ésima observación muestral} \\ \text{más pequeña}\end{matrix}\right)$$

con $i = 1, \ldots, n$. Cada uno de esos pares se grafica como un punto en un sistema de coordenadas bidimensional. Si los percentiles muestrales se acercan a los percentiles de distribución de población correspondientes, el primer número en cada par será aproximadamente igual al segundo número. Los puntos graficados quedarán entonces cerca de una línea a 45°. Desviaciones sustanciales de los puntos graficados con respecto a una línea a 45° hacen dudar de la suposición de que la distribución considerada es la correcta.

Ejemplo 4.29 Un experimentador conoce el valor de cierta constante física. El experimentador realiza $n = 10$ mediciones independientes de este valor por medio de un dispositivo de medición particular y anota los errores de medición resultantes (error = valor observado − valor verdadero). Estas observaciones aparecen en la tabla adjunta.

Porcentaje	5	15	25	35	45
percentil z	−1.645	−1.037	−.675	−.385	−.126
Observación muestral	−1.91	−1.25	−.75	−.53	.20
Porcentaje	55	65	75	85	95
z percentil	.126	.385	.675	1.037	1.645
Observación muestral	.35	.72	.87	1.40	1.56

¿Es factible que el *error de medición* de una variable aleatoria tenga una distribución normal estándar? Los percentiles (z) normales estándar requeridos también se muestran en la tabla. Por lo tanto los puntos en la gráfica de probabilidad son $(-1.645, -1.91)$, $(-1.037, -1.25), \ldots,$ y $(1.645, 1.56)$. La figura 4.33 muestra la gráfica resultante. Aunque los puntos se desvían un poco de la línea a 45°, la impresión predominante es que la línea se

adapta muy bien a los puntos. La gráfica sugiere que la distribución normal estándar es un modelo de probabilidad razonable para el error de medición.

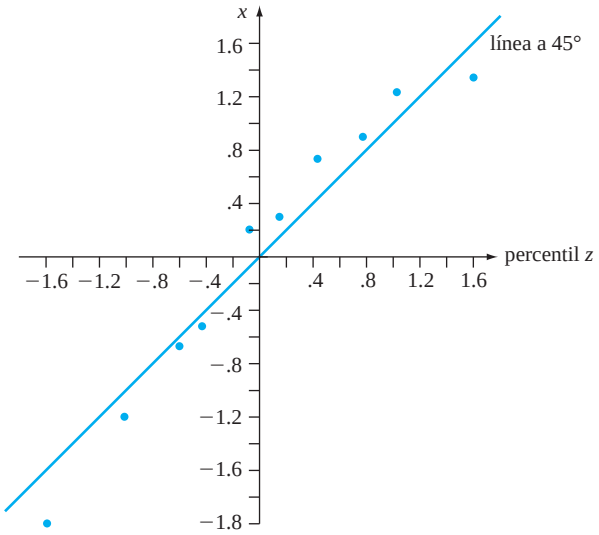


Figura 4.33 Gráficas de pares (percentil z, valor observado) con los datos del ejemplo 4.29; primera muestra

La figura 4.34 muestra una gráfica de pares (percentil z, observación) de una segunda muestra de diez observaciones. La línea a 45° da una buena adaptación a la parte media de la muestra pero no a los extremos. La gráfica tiene apariencia S bien definida. Las dos observaciones muestrales más pequeñas son considerablemente más grandes que los percentiles z correspondientes (los puntos a la extrema izquierda de la gráfica están bien por arriba de la línea a 45°). Asimismo, las dos observaciones muestrales más grandes son mucho más pequeñas que los percentiles z asociados. Esta gráfica indica que la distribución normal estándar no sería una opción factible para el modelo de probabilidad que dio lugar a estos errores de medición observados.

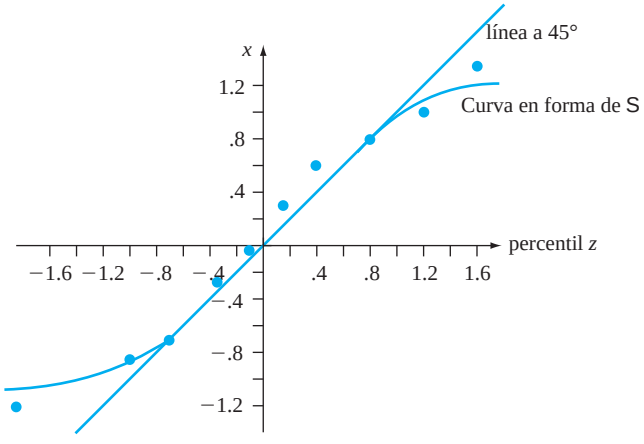


Figura 4.34 Gráficas de pares (percentil z, valor observado) con los datos del ejemplo 4.29; segunda muestra

A un investigador en general no le interesa saber con exactitud si una distribución de probabilidad especificada, tal como la distribución normal estándar (normal con $\mu = 0$ y $\sigma = 1$) o la distribución exponencial con $\lambda = .1$, es un modelo plausible de la distribución de población de la cual se seleccionó la muestra. En cambio, la cuestión es si *algún* miembro de una familia de distribuciones de probabilidad especifica un modelo plausible, la familia de distribuciones normales, la familia de distribuciones exponenciales, la familia de distribuciones Weibull, y así sucesivamente. Los valores de los parámetros de una distribución casi nunca se especifican al principio. Si la familia de distribuciones Weibull se considera como modelo de datos de duración, ¿existen *algunos* valores de los parámetros α y β con los cuales la distribución de Weibull correspondiente se adapte bien a los datos? Afortunadamente, casi siempre es el caso de que sólo una gráfica de probabilidad bastará para evaluar la factibilidad de una familia completa. Si la gráfica se desvía sustancialmente de una línea recta, ningún miembro de la familia es factible. Cuando la gráfica es bastante recta, se requiere más trabajo para estimar valores de los parámetros que generen la distribución más razonable del tipo especificado.

Habrá que enfocarse en una gráfica para verificar la normalidad. Tal gráfica es útil en trabajo aplicado porque muchos procedimientos estadísticos formales dan inferencias precisas sólo cuando la distribución de población es por lo menos aproximadamente normal. Estos procedimientos en general no deben ser utilizados si la gráfica de probabilidad normal muestra un alejamiento muy pronunciado de la linealidad. La clave para construir una gráfica de probabilidad normal que comprenda varios elementos es la relación entre los percentiles (z) normales estándar y aquellos de cualquier otra distribución normal:

$$\text{percentil de una distribución normal } (\mu, \sigma) = \mu + \sigma \cdot (\text{percentil } z \text{ correspondiente})$$

Considérese primero el caso, $\mu = 0$. Si cada observación es exactamente igual al percentil normal correspondiente para algún valor de σ, los pares ($\sigma \cdot$[percentil z], observación) quedan sobre una línea a 45°, cuya pendiente es 1. Esto implica que los pares (percentil z, observación) quedan sobre una recta que pasa por (0, 0) (es decir, una con intersección y en 0) pero con pendiente σ en lugar de 1. El efecto del valor no cero de μ es simplemente cambiar la intersección y de 0 a μ.

Una gráfica de los n pares

$$([100(i - .5)/n]^{\text{o}} \text{ percentil } z, \text{ observación } i\text{-ésima más pequeña})$$

en un sistema de coordenadas bidimensional se llama **gráfica de probabilidad normal.** Si las observaciones muestrales se extraen en realidad de una distribución normal con valor medio μ y desviación estándar σ, los puntos deberán quedar cerca de una línea recta con pendiente σ e intersección en μ. Así pues, una gráfica en la cual los puntos quedan cerca de alguna línea recta sugiere que la suposición de una distribución de población normal es factible.

Ejemplo 4.30 La muestra adjunta compuesta de $n = 20$ observaciones de voltaje de ruptura dieléctrica de un pedazo de resina epóxica apareció en el artículo "Maximum Likelihood Estimation in the 3-Parameter Weibull Distribution" (*IEEE Trans. on Dielectrics and Elec. Insul.*, 1996: 43–55). Los valores de $(i - .5)/n$ para los cuales se requieren los percentiles z son $(1 - .5)/20 = .025$, $(2 - .5)/20 = .075$, . . . , y .975.

Observación	24.46	25.61	26.25	26.42	26.66	27.15	27.31	27.54	27.74	27.94
percentil z	−1.96	−1.44	−1.15	−.93	−.76	−.60	−.45	−.32	−.19	−.06
Observación	27.98	28.04	28.28	28.49	28.50	28.87	29.11	29.13	29.50	30.88
percentil z	.06	.19	.32	.45	.60	.76	.93	1.15	1.44	1.96

La figura 4.35 muestra la gráfica de probabilidad normal resultante. La configuración en la gráfica es bastante recta, lo que indica que es factible que la distribución de la población de voltaje de ruptura dieléctrica sea normal.

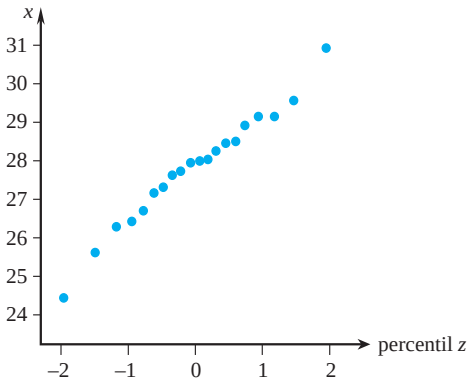


Figura 4.35 Gráfica de probabilidad normal de la muestra de voltaje de ruptura dieléctrica ■

Existe una versión alternativa de una gráfica de probabilidad normal en la cual el eje de los percentiles z es reemplazado por un eje de probabilidad no lineal. La graduación de este eje se construye de modo que los puntos graficados de nuevo queden cerca de una línea cuando la distribución muestreada es normal. La figura 4.36 muestra una gráfica como ésa generada por Minitab con los datos de voltaje de ruptura del ejemplo 4.30.

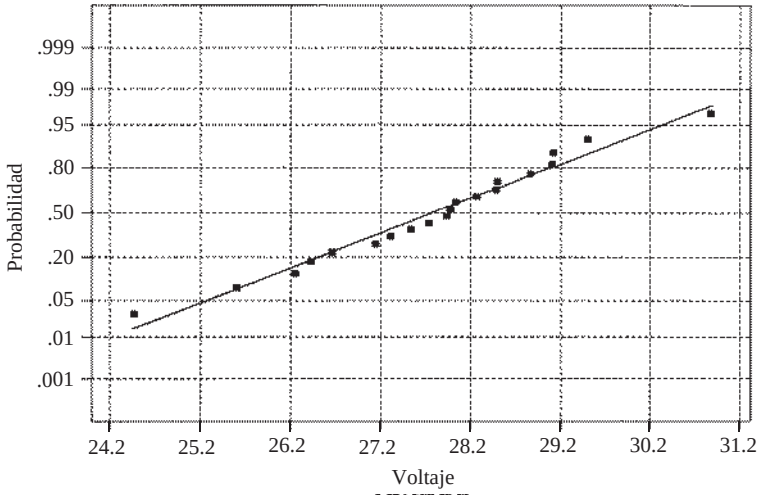


Figura 4.36 Gráfica de probabilidad normal de los datos de voltaje de ruptura generada por Minitab

Una distribución de población no normal a menudo puede ser colocada en una de las siguientes tres categorías:

1. Es simétrica y tiene "colas más livianas" que una distribución normal; es decir, la curva de densidad declina con más rapidez en las colas que una curva normal.
2. Es simétrica y con colas pesadas en comparación con una distribución normal.
3. Es asimétrica.

Una distribución uniforme es de cola liviana, puesto que su función de densidad se reduce a cero afuera de un intervalo finito. La función de densidad $f(x) = 1/[\pi(1 + x^2)]$ para $-\infty < x < \infty$ es de cola pesada, puesto que $1/(1 + x^2)$ declina mucho menos rápidamente que $e^{-x^2/2}$. Las distribuciones lognormal y de Weibull se encuentran entre aquellas que son asimétricas. Cuando los puntos en una gráfica de probabilidad normal no se adhieren a una línea recta, la configuración con frecuencia sugerirá que la distribución de la población se encuentra en una particular de estas tres categorías.

Cuando la distribución de la cual se selecciona la muestra es de cola liviana, las observaciones más grande y más pequeña en general no son tan extremas como podría esperarse de una muestra aleatoria normal. Visualícese una recta trazada a través de la parte media de la gráfica; los puntos a la extrema derecha tienden a estar debajo de la recta (valor observado $<$ percentil z) en tanto que los puntos a la extrema izquierda de la gráfica tienden a quedar sobre la recta (valor observado $>$ percentil z). El resultado es una configuración en forma de S del tipo ilustrado en la figura 4.34.

Una muestra tomada de una distribución de cola pesada también tiende a producir una gráfica en forma de S. Sin embargo, en contraste con el caso de cola liviana, el extremo izquierdo de la gráfica se curva hacia abajo (observado $<$ percentil z), como se muestra en la figura 4.37(a). Si la distribución subyacente es positivamente asimétrica (una cola izquierda corta y una cola derecha larga), las observaciones muestrales más pequeñas serán más grandes que las esperadas de una muestra normal y también lo serán las observaciones más grandes. En este caso, los puntos en ambos extremos de la gráfica quedarán sobre una recta que pasa por la parte media, que produce una configuración curvada, como se ilustra en la figura 4.37(b). Una muestra tomada de una distribución lognormal casi siempre producirá la configuración mencionada. Una gráfica de pares (percentil z, $\ln(x)$) deberá parecerse entonces a una línea recta.

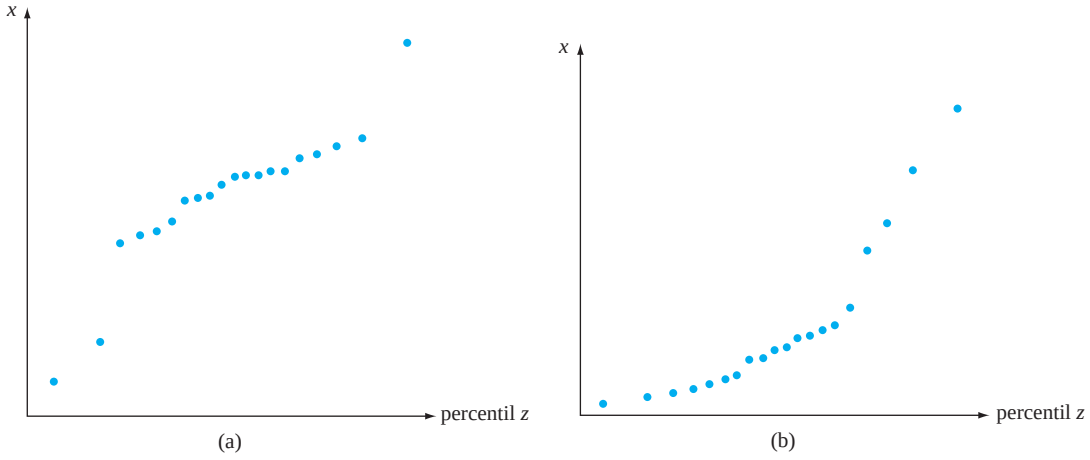


Figura 4.37 Gráficas de probabilidad que sugieren una distribución no normal: (a) una gráfica compatible con una distribución de cola pesada; (b) una gráfica compatible con una distribución positivamente asimétrica

Aunque la distribución de la población sea normal, los percentiles muestrales no coincidirán exactamente con los teóricos debido a la variabilidad del muestreo. ¿Qué tanto pueden desviarse los puntos de la gráfica de probabilidad de un patrón de línea recta antes de que la suposición de normalidad ya no sea plausible? Ésta no es una pregunta fácil de responder. En general, es más probable que una muestra pequeña muestra de una distribución normal produzca una gráfica con un patrón no lineal que una muestra grande. El libro *Fitting Equations to Data* (véase la bibliografía del capítulo 13) presenta los resultados de un estudio de simulación en el cual se seleccionaron numerosas muestras de diferentes tamaños de distribuciones normales. Los autores concluyeron que generalmente varía

mucho la apariencia de la gráfica de probabilidad con tamaños de muestra de menos de 30 y sólo con tamaños de muestra mucho más grandes en general predomina el patrón lineal. Cuando una gráfica está basada en un pequeño tamaño de muestra, sólo un alejamiento muy sustancial de la linealidad se deberá considerar como evidencia concluyente de no normalidad. Un comentario similar se aplica a gráficas de probabilidad para comprobar la factibilidad de otros tipos de distribuciones.

Más allá de la normalidad

Considérese una familia de distribuciones de probabilidad que implica dos parámetros θ_1 y θ_2 y sea $F(x; \theta_1, \theta_2)$ la función de distribución acumulativa correspondiente. La familia de distribuciones normales es una esas familias, con $\theta_1 = \mu$, $\theta_2 = \sigma$ y $F(x; \mu, \sigma) = \Phi[(x - \mu)/\sigma]$. Otro ejemplo es la familia de Weibull, con $\theta_1 = \alpha$, $\theta_2 = \beta$ y

$$F(x; \alpha, \beta) = 1 - e^{-(x/\beta)^\alpha}$$

Otra familia más de este tipo es la familia gamma, para la cual la función de distribución acumulativa es una integral que implica la función gamma incompleta que no puede ser expresada en alguna forma más simple.

Se dice que los parámetros θ_1 y θ_2 son **parámetros de ubicación** y **escala,** respectivamente, si $F(x; \theta_1, \theta_2)$ es una función de $(x - \theta_1)/\theta_2$. Los parámetros μ y σ de la familia normal son los parámetros de ubicación y escala, respectivamente. Al cambiar μ, la curva de densidad en forma de campana se desplaza a la derecha o izquierda y al cambiar σ se alarga o comprime la escala de medición (la escala sobre el eje horizontal cuando se grafica la función de densidad). La función de distribución acumulativa da otro ejemplo

$$F(x; \theta_1, \theta_2) = 1 - e^{-e^{(x-\theta_1)/\theta_2}} \quad -\infty < x < \infty$$

Se dice que una variable aleatoria con esta función de distribución acumulativa tiene una *distribución de valor extremo*. Se utiliza en aplicaciones que implican la duración de un componente y la resistencia de un material.

Aunque la forma de la función de distribución acumulativa de valor extremo a primera vista pudiera sugerir que θ_1 es el punto de simetría de la función de densidad y por ende la media y la mediana, éste no es el caso. En cambio, $P(X \le \theta_1) = F(\theta_1; \theta_1, \theta_2) = 1 - e^{-1} = .632$ y la función de densidad $f(x; \theta_1, \theta_2) = F'(x; \theta_1, \theta_2)$ es negativamente asimétrica (una larga cola inferior). Asimismo, el parámetro de escala θ_2 no es la desviación estándar ($\mu = \theta_1 - .5772\theta_2$ y $\sigma = 1.283\theta_2$). Sin embargo, al cambiar el valor de θ_1 cambia la ubicación de la curva de densidad, mientras que al cambiar θ_2 cambia la escala del eje de medición.

El parámetro β de la distribución de Weibull es un parámetro de escala, pero α no es un parámetro de ubicación. Un comentario similar es pertinente para los parámetros α y β de la distribución gamma. En la forma usual, la función de densidad de cualquier miembro de la distribución gamma o de Weibull es positiva para $x > 0$ y cero de otra manera. Un parámetro de ubicación puede ser introducido como tercer parámetro γ (se hizo esto para la distribución de Weibull) para desplazar la función de densidad de modo que sea positiva si $x > \gamma$ y cero de otra manera.

Cuando la familia considerada tiene sólo parámetros de ubicación y escala, el tema de si cualquier miembro de la familia es una distribución de población plausible puede ser abordado vía una gráfica de probabilidad única de fácil construcción. Primero se obtienen los percentiles de la *distribución estándar*, una con $\theta_1 = 0$ y $\theta_2 = 1$, con los porcentajes $100(i - .5)/n$ $(i = 1, \ldots, n)$. Los n pares (percentil estandarizado, observación) dan los puntos en la gráfica. Esto es exactamente lo que se hizo para obtener una gráfica de probabilidad normal ómnibus. Un tanto sorprendentemente, esta metodología puede ser aplicada para dar una gráfica de probabilidad Weibull más general. El resultado clave es que si X tiene una distribución de Weibull con parámetro de forma α y parámetro de escala β, entonces la variable transformada $\ln(X)$ tiene una distribución de valor extremo con parámetro de ubicación $\theta_1 = \ln(\beta)$ y parámetro de escala $1/\alpha$. Así pues una gráfica de los

pares (percentil estandarizado de valor extremo, ln(x)) que muestre un fuerte patrón lineal apoya la selección de la distribución de Weibull como modelo de una población.

Ejemplo 4.31 Las observaciones adjuntas son de la duración (en horas) del aislamiento de aparatos eléctricos cuando la aceleración del esfuerzo térmico y eléctrico se mantuvo fija en valores particulares ("On the Estimation of Life of Power Apparatus Insulation Under Combined Electrical and Thermal Stress", *IEEE Trans. on Electrical Insulation*, 1985: 70–78). Una gráfica de probabilidad de Weibull necesita calcular primero los percentiles 5º, 15º, . . . , y 95º de la distribución de valor extremo estándar. El $(100p)$º percentil $\eta(p)$ satisface

$$p = F(\eta(p)) = 1 - e^{-e^{\eta(p)}}$$

de donde $\eta(p) = \ln[-\ln(1 - p)]$.

Percentil	−2.97	−1.82	−1.25	−.84	−.51
x	282	501	741	851	1072
$ln(x)$	5.64	6.22	6.61	6.75	6.98
Percentil	−.23	.05	.33	.64	1.10
x	1122	1202	1585	1905	2138
$ln(x)$	7.02	7.09	7.37	7.55	7.67

Los pares (−2.97, 5.64), (−1.82, 6.22), . . . , (1.10, 7.67) se grafican como puntos en la figura 4.38. La derechura de la gráfica argumenta firmemente a favor del uso de la distribución de Weibull como modelo de duración del aislamiento, una conclusión también alcanzada por el autor del citado artículo.

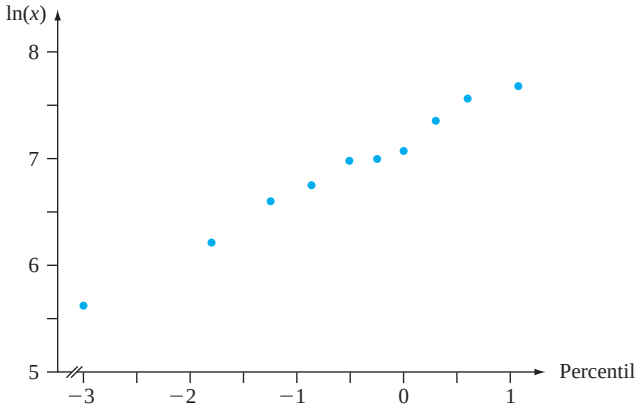


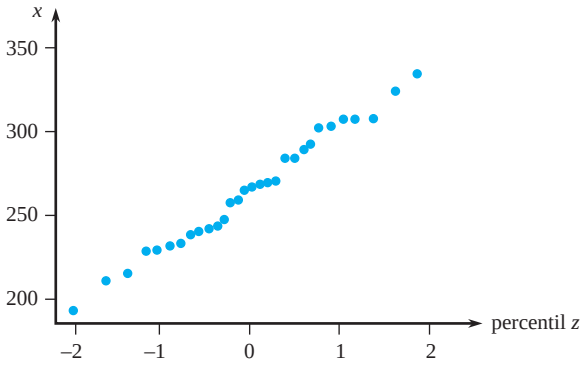
Figura 4.38 Gráfica de probabilidad Weibull de los datos de duración del aislamiento ■

La distribución gamma es un ejemplo de una familia que implica un parámetro de forma para el cual no hay ninguna transformación $h(\cdot)$ de tal suerte que $h(X)$ tenga una distribución que dependa sólo de los parámetros de ubicación y escala. Para construir una gráfica de probabilidad primero se tiene que estimar el parámetro de forma de los datos muestrales (algunos métodos para realizar lo anterior se describen en el capítulo 6). En ocasiones un investigador desea saber si la variable transformada X^θ tiene una distri-

bución normal para algún valor de θ (por convención, $\theta = 0$ es idéntica a la transformación, logarítmica en cuyo caso X tiene una distribución lognormal). El libro *Graphical Methods for Data Analysis*, citado en la bibliografía del capítulo 1, discute este tipo de problema así como también otros refinamientos de construcción de gráficas de probabilidad. Afortunadamente, la amplia disponibilidad de varias gráficas de probabilidad junto con paquetes de software estadísticos significa que el usuario con frecuencia puede evitar los detalles técnicos.

EJERCICIOS Sección 4.6 (87–97)

87. La gráfica de probabilidad normal adjunta se construyó con una muestra de 30 lecturas de tensión de pantallas de malla localizadas detrás de la superficie de tubos de pantallas de video utilizadas en monitores de computadora. ¿Parece factible que la distribución de tensión sea normal?



88. Una muestra de 15 golfistas universitarias fue seleccionada y la velocidad de la cabeza del palo (km/h) mientras se hace *swing* con un *driver* se determinó para cada una, dando como resultado los siguientes datos ("Hip Rotational Velocities During the Full Golf Swing", *J. of Sports Science and Medicine*, 2009: 296–299):

69.0	69.7	72.7	80.3	81.0
85.0	86.0	86.3	86.7	87.7
89.3	90.7	91.0	92.5	93.0

Los correspondientes percentiles z son

−1.83	−1.28	−0.97	−0.73	−0.52
−0.34	−0.17	0.0	0.17	0.34
0.52	0.73	0.97	1.28	1.83

Construya una gráfica de probabilidad normal y una gráfica de puntos. ¿Es factible que la distribución de la población sea normal?

89. Construya una gráfica de probabilidad normal con la siguiente muestra de observaciones de espesor de recubrimiento de pintura de baja viscosidad ("Achieving a Target Value for a Manufacturing Process: A Case Study", *J. of Quality Technology*, 1992: 22–26). ¿Se sentiría cómodo estimando el espesor medio de la población con un método que supuso una distribución de población normal?

.83	.88	.88	1.04	1.09	1.12	1.29	1.31
1.48	1.49	1.59	1.62	1.65	1.71	1.76	1.83

90. El artículo "A Probabilistic Model of Fracture in Concrete and Size Effects on Fracture Toughness" (*Magazine of Concrete Res.*, 1996: 311–320) da argumentos de por qué la distribución de tenacidad a la fractura en especímenes de concreto debe tener una distribución de Weibull y presenta varios histogramas de datos a los que adaptan bien curvas de Weibull superpuestas. Considere la siguiente muestra de tamaño $n = 18$ observaciones de tenacidad de concreto de alta resistencia (compatible con uno de los histogramas); también se dan los valores de $p_i = (i - .5)/18$.

Observación	.47	.58	.65	.69	.72	.74
p_i	.0278	.0833	.1389	.1944	.2500	.3056
Observación	.77	.79	.80	.81	.82	.84
p_i	.3611	.4167	.4722	.5278	.5833	.6389
Observación	.86	.89	.91	.95	1.01	1.04
p_i	.6944	.7500	.8056	.8611	.9167	.9722

Construya una gráfica de probabilidad de Weibull y coméntela.

91. Construya una gráfica de probabilidad normal con los datos de propagación de grietas por fatiga dados en el ejercicio 39 (capítulo 1). ¿Parece factible que la duración de la propagación tenga una distribución normal? Explique.

92. El artículo "The Load-Life Relationship for M50 Bearings with Silicon Nitride Ceramic Balls" (*Lubrication Engr.*, 1984: 153–159) reporta los datos adjuntos de duración de cojinetes (millones de revoluciones) probados con una carga de 6.45 kN.

47.1	68.1	68.1	90.8	103.6	106.0	115.0
126.0	146.6	229.0	240.0	240.0	278.0	278.0
289.0	289.0	367.0	385.9	392.0	505.0	

a. Construya una gráfica de probabilidad normal. ¿La normalidad es plausible?

b. Construya una gráfica de probabilidad de Weibull. ¿Es adecuada la familia de distribución de Weibull?

93. Construya una gráfica de probabilidad que le permita evaluar qué tan adecuada es la distribución lognormal como modelo de los datos de cantidad de lluvia del ejercicio 83 (capítulo 1).

94. Las observaciones adjuntas son valores de precipitación durante marzo a lo largo de un periodo de 30 años en Minneapolis-St. Paul.

.77	1.20	3.00	1.62	2.81	2.48
1.74	.47	3.09	1.31	1.87	.96
.81	1.43	1.51	.32	1.18	1.89
1.20	3.37	2.10	.59	1.35	.90
1.95	2.20	.52	.81	4.75	2.05

a. Construya e interprete una gráfica de probabilidad normal con este conjunto de datos.
b. Calcule la raíz cuadrada de cada valor y luego construya una gráfica de probabilidad normal basada en estos datos transformados. ¿Parece factible que la raíz cuadrada de la precipitación esté normalmente distribuida?
c. Repita el inciso (b) después de transformar por medio de raíces cúbicas.

95. Use un paquete de software estadístico para construir una gráfica de probabilidad normal de los datos de resistencia última a la tensión dados en el ejercicio 13 del capítulo 1 y comente.

96. Sean $y_1, y_2, \ldots, y_n$, las observaciones muestrales *ordenadas* (con y_1 como la más pequeña y y_n como la más grande). Una verificación sugerida de normalidad es graficar los pares $(\Phi^{-1}((i - .5)/n), y_i)$. Suponga que se cree que las observaciones provienen de una distribución con media 0 y sean $w_1, \ldots, w_n$ los valores *absolutos ordenados* de las x_i. Una gráfica **seminormal** es una gráfica de probabilidad de las w_i. Más específicamente, como $P(|Z| \le w) = P(-w \le Z \le w) = 2\Phi(w) - 1$, una gráfica seminormal es una gráfica de los pares $(\Phi^{-1}\{[(i - .5)/n + 1]/2\}, w_i)$. La virtud de ésta es que los valores apartados pequeños o grandes en la muestra original ahora aparecerán sólo en el extremo superior de la gráfica y no en ambos extremos. Construya una gráfica seminormal con la siguiente muestra de errores de medición, y comente: −3.78, −1.27, 1.44, −.39, 12.38, −43.40, 1.15, −3.96, −2.34, 30.84.

97. Las siguientes observaciones de tiempo de falla (miles de horas) se obtuvieron con una prueba de duración acelerada de 16 chips de circuitos integrados de un tipo:

82.8	11.6	359.5	502.5	307.8	179.7
242.0	26.5	244.8	304.3	379.1	212.6
229.9	558.9	366.7	204.6		

Use los percentiles correspondientes de la distribución exponencial con $\lambda = 1$ para construir una gráfica de probabilidad. Luego explique por qué la gráfica valora la aptitud de la muestra habiendo sido generada a partir de *cualquier* distribución exponencial.

EJERCICIOS SUPLEMENTARIOS (98–128)

98. Sea X = el tiempo que una cabeza de lectura-escritura requiere para localizar un registro deseado en un dispositivo de memoria de disco de computadora una vez que la cabeza se ha colocado sobre la pista correcta. Si los discos giran una vez cada 25 milisegundos, una suposición razonable es que X está uniformemente distribuida en el intervalo [0, 25].
a. Calcule $P(10 \le X \le 20)$.
b. Calcule $P(X \ge 10)$.
c. Obtenga la función de distribución acumulativa $F(X)$.
d. Calcule $E(X)$ y σ_X.

99. Una barra de 12 pulg que está sujeta por ambos extremos se somete a una cantidad creciente de esfuerzo hasta que se rompe. Sea Y = la distancia del extremo izquierdo al punto donde ocurre la ruptura. Suponga que Y tiene la función de densidad de probabilidad

$$f(y) = \begin{cases} \left(\frac{1}{24}\right) y \left(1 - \frac{y}{12}\right) & 0 \le y \le 12 \\ 0 & \text{de lo contrario} \end{cases}$$

Calcule lo siguiente:
a. La función de distribución acumulativa de Y y grafíquela.
b. $P(Y \le 4)$, $P(Y > 6)$ y $P(4 \le Y \le 6)$
c. $E(Y)$, $E(Y^2)$ y $V(Y)$.
d. La probabilidad de que el punto de ruptura ocurra a más de 2 pulg del punto de ruptura esperado.
e. La longitud esperada del segmento más corto cuando ocurre la ruptura.

100. Sea X el tiempo hasta la falla (en años) de cierto componente hidráulico. Suponga que la función de densidad de probabilidad de X es $f(x) = 32/(x + 4)^3$ con $x > 0$.
a. Verifique que $f(x)$ es una función de densidad de probabilidad legítima.
b. Determine la función de distribución acumulativa.
c. Use el resultado del inciso (b) para calcular la probabilidad de que el tiempo hasta la falla sea de entre 2 y 5 años.
d. ¿Cuál es el tiempo esperado hasta la falla?
e. Si el componente tiene un valor de recuperación igual a $100/(4 + x)$ cuando su tiempo hasta la falla es x, ¿cuál es el valor de recuperación esperado?

101. El tiempo X para la terminación de cierta tarea tiene una función de distribución acumulativa $F(x)$ dada por

$$\begin{cases} 0 & x < 0 \\ \frac{x^3}{3} & 0 \le x < 1 \\ 1 - \frac{1}{2}\left(\frac{7}{3} - x\right)\left(\frac{7}{4} - \frac{3}{4}x\right) & 1 \le x \le \frac{7}{3} \\ 1 & x > \frac{7}{3} \end{cases}$$

a. Obtenga la función de densidad de probabilidad $f(x)$ y trace su gráfica.
b. Calcule $P(.5 \le X \le 2)$.
c. Calcule $E(X)$.

102. Se sabe que el voltaje de ruptura de un diodo seleccionado al azar de cierto tipo está normalmente distribuido con valor medio de 40 V y desviación estándar de 1.5 V.
- **a.** ¿Cuál es la probabilidad de que el voltaje de un solo diodo sea de entre 39 y 42?
- **b.** ¿Qué valor es tal que sólo 15% de todos los diodos tienen voltajes que excedan dicho valor?
- **c.** Si se seleccionan cuatro diodos independientemente, ¿cuál es la probabilidad de que por lo menos uno tenga un voltaje de más de 42?

103. El artículo "Computer Assisted Net Weight Control" (*Quality Progress*, 1983: 22–25) sugiere una distribución normal con media de 137.2 oz y desviación estándar de 1.6 oz del contenido real de frascos de cierto tipo. El contenido declarado fue de 135 oz.
- **a.** ¿Cuál es la probabilidad de que un solo frasco contenga más que el contenido declarado?
- **b.** Entre diez frascos seleccionados al azar, ¿cuál es la probabilidad de que por lo menos ocho contengan más que el contenido declarado?
- **c.** Suponiendo que la media permanece en 137.2, ¿a qué valor se tendría que cambiar la desviación estándar de modo que 95% de todos los frascos contengan más que el contenido declarado?

104. Cuando tarjetas de circuito utilizadas en la fabricación de reproductores de discos compactos se someten a prueba, el porcentaje a largo plazo de tarjetas defectuosas es de 5%. Suponga que se recibió un lote de 250 tarjetas y que la condición de cualquier tarjeta particular es independiente de la de cualquier otra.
- **a.** ¿Cuál es la probabilidad aproximada de que por lo menos 10% de las tarjetas en el lote sean defectuosas?
- **b.** ¿Cuál es la probabilidad aproximada de que haya exactamente 10 defectuosas en el lote?

105. El artículo "Characterization of Room Temperature Damping in Aluminum-Indium Alloys" (*Metallurgical Trans.*, 1993: 1611–1619) sugiere que el tamaño de grano de matriz A1 (μm) de una aleación compuesta de 2% de indio podría ser modelado con una distribución normal con valor medio de 96 y desviación estándar de 14.
- **a.** ¿Cuál es la probabilidad de que el tamaño de grano exceda de 100?
- **b.** ¿Cuál es la probabilidad de que el tamaño de grano sea de entre 50 y 80?
- **c.** ¿Qué intervalo (a, b) incluye el 90% central de todos los tamaños de grano (de modo que 5% esté por debajo de a y 5% por encima de b)?

106. El tiempo de reacción (en segundos) a un estímulo es una variable aleatoria continua con función de densidad de probabilidad

$$f(x) = \begin{cases} \frac{3}{2} \cdot \frac{1}{x^2} & 1 \le x \le 3 \\ 0 & \text{de lo contrario} \end{cases}$$

- **a.** Obtenga la función de distribución acumulativa.
- **b.** ¿Cuál es la probabilidad de que el tiempo de reacción sea cuando mucho de 2.5 s? ¿De entre 1.5 y 2.5 s?
- **c.** Calcule el tiempo de reacción esperado.
- **d.** Calcule la desviación estándar del tiempo de reacción.
- **e.** Si un individuo requiere de más de 1.5 s para reaccionar, una luz se enciende y permanece encendida hasta que transcurre un segundo más o hasta que la persona reacciona (lo que suceda primero). Determine la cantidad de tiempo que se espera que la luz permanezca encendida. [*Sugerencia*: sea $h(X)$ = el tiempo que la luz está encendida como una función del tiempo de reacción X.]

107. Sea X la temperatura a la cual ocurre una reacción química. Suponga que X tiene una función de densidad de probabilidad

$$f(x) = \begin{cases} \frac{1}{9}(4 - x^2) & -1 \le x \le 2 \\ 0 & \text{de lo contrario} \end{cases}$$

- **a.** Trace la gráfica de $f(x)$.
- **b.** Determine la función de distribución acumulativa y grafíquela.
- **c.** ¿Es 0 la temperatura mediana a la cual ocurre la reacción? Si no, ¿es la temperatura mediana más pequeña o más grande que 0?
- **d.** Suponga que esta reacción se realiza independientemente una vez en cada uno de diez laboratorios diferentes y que la función de densidad de probabilidad del tiempo de reacción en cada laboratorio es como se da. Sea Y = el número entre los diez laboratorios en los cuales la temperatura excede de 1. ¿Qué clase de distribución tiene Y? (Dé los nombres y valores de los parámetros.)

108. El artículo "Determination of the MTF of Positive Photoresists Using the Monte Carlo Method" (*Photographic Sci. and Engr.*, 1983: 254–260) propone la distribución exponencial con parámetro $\lambda = .93$ como modelo de la distribución de una longitud de trayectoria libre de fotones (μm) en ciertas circunstancias. Suponga que éste es el modelo correcto.
- **a.** ¿Cuál es la longitud de trayectoria esperada y cuál es la desviación estándar de ésta?
- **b.** ¿Cuál es la probabilidad de que la longitud de trayectoria exceda de 3.0? ¿Cuál es la probabilidad de que la longitud de trayectoria esté entre 1.0 y 3.0?
- **c.** ¿Qué valor es excedido por sólo 10% de todas las longitudes de trayectoria?

109. El artículo "The Prediction of Corrosion by Statistical Analysis of Corrosion Profiles" (*Corrosion Science*, 1985: 305–315) sugiere la siguiente función de distribución acumulativa de la profundidad X del pozo más profundo en un experimento que implica la exposición de acero al carbono manganeso a agua de mar acidificada.

$$F(x; \alpha, \beta) = e^{-e^{-(x-\alpha)/\beta}} \quad -\infty < x < \infty$$

Los autores proponen los valores $\alpha = 150$ y $\beta = 90$. Suponga que éste es el modelo correcto.
- **a.** ¿Cuál es la probabilidad de que la profundidad del pozo más profundo sea cuando mucho de 150? ¿Cuando mucho 300? ¿De entre 150 y 300?
- **b.** ¿Por debajo de qué valor estará la profundidad del pozo máximo en 90% de todos los experimentos?
- **c.** ¿Cuál es la función de densidad de X?
- **d.** Se puede demostrar que la función de densidad es unimodal (una sola cresta). ¿Por encima de qué valor sobre el eje de medición ocurre esta cresta? (Este valor es la moda.)

e. Se puede demostrar que $E(X) \approx .5772\beta + \alpha$. ¿Cuál es la media de los valores dados de α y β y cómo se compara con la mediana y la moda? Trace la gráfica de la función de densidad. [*Nota*: ésta se conoce como *distribución de valor extremo más grande*.]

110. Sea t = la cantidad de impuesto sobre las ventas que un minorista debe al gobierno por un periodo determinado. El artículo "Statistical Sampling in Tax Audits" (*Statistics and the Law*, 2008: 320–343) se propone modelar la incertidumbre en t, considerándola como una variable aleatoria distribuida normalmente con media μ y la desviación estándar σ (en el artículo, estos dos parámetros se estiman a partir de los resultados de una inspección fiscal que implican n operaciones de muestreo). Si a representa la cantidad a la que el minorista es evaluado, entonces resulta una subevaluación si $t > a$ y una sobreevaluación de resultados, si $a > t$. La función de sanción propuesta (es decir, la pérdida) para la sobre o subevaluación es $L(a, t) = t - a$ si $t > a$ y $= k(a - t)$ si $t \le a$ ($k > 1$ se sugiere para incorporar la idea de que la sobreevaluación es más grave que una subevaluación).

a. Demuestre que $a^* = \mu + \sigma\Phi^{-1}(1/(k + 1))$ es el valor de a que minimiza la pérdida esperada, donde Φ^{-1} es la función inversa de la función de distribución acumulativa normal estándar.

b. Si $k = 2$ (se sugiere en el artículo), $\mu =$ \$100 000, y $\sigma =$ \$10 000, ¿cuál es el valor óptimo de a, y cuál es la probabilidad resultante de la sobreevaluación?

111. La *moda* de una distribución continua es el valor x^* que incrementa al máximo $f(x)$.

a. ¿Cuál es la moda de una distribución normal con parámetros μ y σ?

b. ¿Tiene una sola moda la distribución uniforme con parámetros A y B? ¿Por qué sí o por qué no?

c. ¿Cuál es la moda de una distribución exponencial con parámetro λ? (Trace una gráfica.)

d. Si X tiene una distribución gamma con parámetros α y β, y $\alpha > 1$, halle la moda. [*Sugerencia*: $\ln[f(x)]$ se incrementará al máximo si y sólo si $f(x)$ es, y puede ser más simple sacar la derivada de $\ln[f(x)]$.

e. ¿Cuál es la moda de una distribución ji cuadrada con ν grados de libertad?

112. El artículo "Error Distribution in Navigation" (*J. of the Institute of Navigation*, 1971: 429–442) sugiere que una distribución de frecuencia de errores positivos (magnitudes de errores) es mejor aproximada por una distribución exponencial. Sea X = el error de posición lateral (millas náuticas), el cual puede ser positivo o negativo. Suponga que la función de densidad de probabilidad de X es

$$f(x) = (.1)e^{-.2|x|} \quad -\infty < x < \infty$$

a. Trace una gráfica de $f(x)$ y compruebe que $f(x)$ es una función de densidad de probabilidad legítima (demuestre que se integra a 1).

b. Obtenga la función de distribución acumulativa de X y trácela.

c. Calcule $P(X \le 0)$, $P(X \le 2)$, $P(-1 \le X \le 2)$ y la probabilidad de que se cometa un error de más de 2 millas.

113. En algunos sistemas, un cliente es asignado a una o dos prestadoras de servicios. Si el tiempo para que el cliente sea atendido por la prestadora de servicios i tiene una distribución exponencial con parámetro λ_i ($i = 1, 2$) y p es la proporción de todos los clientes atendidos por la prestadora de servicios 1, entonces la función de densidad de probabilidad de X = el tiempo para ser atendido de un cliente seleccionado al azar es

$$f(x; \lambda_1, \lambda_2, p) = \begin{cases} p\lambda_1 e^{-\lambda_1 x} + (1 - p)\lambda_2 e^{-\lambda_2 x} & x \ge 0 \\ 0 & \text{de lo contrario} \end{cases}$$

Ésta a menudo se llama distribución hiperexponencial o distribución exponencial combinada. Esta distribución también se propone como modelo de la cantidad de lluvia en "Modeling Monsoon Affected Rainfall of Pakistan by Point Processes" (*J. of Water Resources Planning and Mgmnt.*, 1992: 671-688).

a. Verifique que $f(x; \lambda_1, \lambda_2, p)$ sí es una función de densidad de probabilidad.

b. ¿Cuál es la función de distribución acumulativa $F(x; \lambda_1, \lambda_2, p)$?

c. Si $f(x; \lambda_1, \lambda_2, p)$ es la función de densidad de probabilidad de X, ¿cuál es $E(X)$?

d. Utilizando el hecho de que $E(X^2) = 2/\lambda^2$ cuando X tiene una distribución exponencial con parámetro λ, calcule $E(X^2)$ cuando X tiene la función de densidad de probabilidad $f(x; \lambda_1, \lambda_2, p)$. Luego calcule $V(X)$.

e. El coeficiente de variación de una variable aleatoria (o distribución) es $CV = \sigma/\mu$. ¿Cuál es CV para una variable aleatoria exponencial? ¿Qué puede decir sobre el valor de CV cuando X tiene una distribución hiperexponencial?

f. ¿Cuál es el CV de una distribución Erlang con parámetros λ y n como se definen en el ejercicio 68? [*Nota*: en trabajo aplicado, el CV muestral se utiliza para decidir cuál de las tres distribuciones podría ser apropiada.]

114. Suponga que en un estado particular se permite que los personas físicas que presentan su declaración de impuestos detallen sus deducciones sólo si el total de las deducciones detalladas es por lo menos de \$5000. Sea X (en miles de dólares) el total de deducciones detalladas en un formulario seleccionado al azar. Suponga que X tiene la función de densidad de probabilidad

$$f(x; \alpha) = \begin{cases} k/x^\alpha & x \ge 5 \\ 0 & \text{de lo contrario} \end{cases}$$

a. Encuentre el valor de k. ¿Qué restricción en α es necesaria?

b. ¿Cuál es la función de distribución acumulativa de X?

c. ¿Cuál es la deducción total esperada en un formulario seleccionado al azar? ¿Qué restricción es necesaria en α para que $E(X)$ sea finita?

d. Demuestre que $\ln(X/5)$ tiene una distribución exponencial con parámetro $\alpha - 1$.

115. Sea I_i la corriente de entrada a un transistor e I_0 la corriente de salida. En ese caso la ganancia de corriente es proporcional a $\ln(I_0/I_i)$. Suponga que la constante de proporcionalidad es 1 (lo que conduce a seleccionar una unidad de medición particular), así que la ganancia de corriente $= X = \ln(I_0/I_i)$. Suponga que X está normalmente distribuida con $\mu = 1$ y $\sigma = .05$.

a. ¿Qué tipo de distribución tiene la razón I_0/I_i?

b. ¿Cuál es la probabilidad de que la corriente de salida sea más de dos veces la corriente de entrada?

c. ¿Cuáles son el valor esperado y la varianza de la razón entre corriente de salida y corriente de entrada?

116. El artículo "Response of SiC_f/Si_3N_4 Composites Under Static and Cyclic Loading–An Experimental and Statistical Analysis" (*J. of Engr. Materials and Technology*, 1997: 186–193) sugiere que la resistencia a la tensión (MPa) de compuestos en condiciones especificadas puede ser modelada por una distribución de Weibull con $\alpha = 9$ y $\beta = 180$.

a. Trace una gráfica de la función de densidad.

b. ¿Cuál es la probabilidad de que la resistencia de un espécimen seleccionado al azar exceda de 175? ¿Sea de entre 150 y 175?

c. Si se seleccionan al azar dos especímenes y sus resistencias son independientes entre sí, ¿cuál es la probabilidad de que por lo menos uno tenga una resistencia de entre 150 y 175?

d. ¿Qué valor de resistencia separa al 10% de todos los especímenes más débiles del 90% restante?

117. Si Z tiene una distribución normal estándar, defina una nueva variable aleatoria Y como $Y = \sigma Z + \mu$. Demuestre que Y tiene una distribución normal con parámetros μ y σ. [*Sugerencia*: $Y \le y$ si y sólo si $Z \le$? Use esto para definir la función de distribución acumulativa de Y y luego diferénciela con respecto a y.]

118. **a.** Suponga que la duración X de un componente, medida en horas, tiene una distribución gamma con parámetros α y β. Sea $Y =$ la duración medida en minutos. Deduzca la función de densidad de probabilidad de Y. [*Sugerencia*: $Y \le y$ si y sólo si $X \le y/60$. Use esto para obtener la función de distribución acumulativa de Y y luego diferénciela para obtener la función de densidad de probabilidad.]

b. Si X tiene una distribución gamma con parámetros α y β, ¿cuál es la distribución de probabilidad de $Y = cX$?

119. En los ejercicios 117 y 118, así como también en muchas otras situaciones, se tiene la función de densidad de probabilidad $f(x)$ de X y se desea conocer la función de densidad de probabilidad de $y = h(x)$. Suponga que $h(\cdot)$ es un función invertible, de modo que $y = h(x)$ se resuelve para x a fin de obtener $x = k(y)$. Entonces se puede demostrar que la función de densidad de probabilidad de Y es

$$g(y) = f[k(y)] \cdot |k'(y)|$$

a. Si X tiene una distribución uniforme con $A = 0$ y $B = 1$, obtenga la función de densidad de probabilidad de $Y = -\ln(X)$.

b. Resuelva el ejercicio 117 utilizando este resultado.

c. Resuelva el ejercicio 118(b) utilizando este resultado.

120. Basado en los datos del experimento de lanzamiento de dardo, el artículo "Shooting Darts" (*Chance*, verano de 1997: 16–19) propuso que los errores horizontales y verticales al apuntar a un blanco deben ser independientes unos de otros, cada uno con una distribución normal con media 0 y varianza σ^2. Se puede demostrar entonces que la distancia V del blanco al punto de aterrizaje es

$$f(v) = \frac{v}{\sigma^2} \cdot e^{-v^2/2\sigma^2} \quad v > 0$$

a. ¿De qué familia introducida en este capítulo es miembro esta función de densidad de probabilidad?

b. Si $\sigma = 20$ mm (cerca del valor sugerido por el artículo), ¿cuál es la probabilidad de que un dardo aterrice dentro de 25 mm (aproximadamente 1 pulg) del blanco?

121. El artículo "Three Sisters Give Birth on the Same Day" (*Chance*, primavera de 2001, 23–25) utilizó el hecho de que tres hermanas de Utah dieron a luz el 11 de marzo de 1998 como base para plantear algunas preguntas interesantes con respecto a coincidencias de fechas de nacimiento.

a. No haciendo caso del año bisiesto y suponiendo que los otros 365 días son igualmente probables, ¿cuál es la probabilidad de que tres nacimientos seleccionados al azar ocurran el 11 de marzo? Asegúrese de indicar qué suposiciones adicionales está haciendo.

b. Con las suposiciones utilizadas en el inciso (a), ¿cuál es la probabilidad de que tres nacimientos seleccionados al azar ocurran el mismo día?

c. El autor sugirió, basado en datos extensos, que el tiempo de gestación (tiempo entre la concepción y el nacimiento) podía ser modelado como si tuviera una distribución normal con valor medio de 280 días y desviación estándar de 19.88 días. Las fechas esperadas para las tres hermanas de Utah fueron el 15 de marzo, el 1 de abril y el 4 de abril, respectivamente. Suponiendo que las tres fechas esperadas están en la media de la distribución, ¿cuál es la probabilidad de que los nacimientos ocurrieran el 11 de marzo? [*Sugerencia*: la desviación de la fecha de nacimiento con respecto a la fecha esperada está normalmente distribuida con media 0.]

d. Explique cómo utilizaría la información del inciso (c) para calcular la probabilidad de una fecha de nacimiento común.

122. Sea X la duración de un componente, con $f(x)$ y $F(x)$ como la función de densidad de probabilidad y la función de distribución acumulativa de X. La probabilidad de que el componente falle en el intervalo $(x, x + \Delta x)$ es aproximadamente $f(x) \cdot \Delta x$. La probabilidad condicional de que falle en $(x, x + \Delta x)$ dado que ha durado por lo menos x es $f(x) \cdot \Delta x/[1 - F(x)]$. Dividiendo esto entre Δx se produce la **función de tasa de falla:**

$$r(x) = \frac{f(x)}{1 - F(x)}$$

Una función de tasa de falla creciente indica que la probabilidad de que los componentes viejos se desgasten es cada vez más grande, mientras que una tasa de falla decreciente evidencia una confiabilidad cada vez más grande con la edad. En la práctica, a menudo se supone una falla "en forma de tina de baño".

a. Si X está exponencialmente distribuida, ¿cuál es $r(x)$?

b. Si X tiene una distribución de Weibull con parámetros α y β, ¿cuál es $r(x)$? ¿Con qué valores de parámetros se incrementará $r(x)$? ¿Con qué valores de parámetros decrecerá $r(x)$ con x?

c. Como $r(x) = -(d/dx)\ln[1 - F(x)]$, $1n[1 - F(x)] = -\int r(x)dx$. Suponga

$$r(x) = \begin{cases} \alpha\left(1 - \dfrac{x}{\beta}\right) & 0 \le x \le \beta \\ 0 & \text{de lo contrario} \end{cases}$$

de modo que si un componente dura β horas, durará por siempre (si bien parece irrazonable, este modelo puede ser utilizado para estudiar el "desgaste inicial"). ¿Cuáles son la función de distribución acumulativa y la función de densidad de probabilidad de X?

123. Sea U que tiene una distribución uniforme en el intervalo [0, 1]. Entonces los valores observados que tienen esta distribución se pueden obtener con un generador de números aleatorios por computadora. Sea $X = -(1/\lambda)\ln(1 - U)$.

a. Demuestre que X tiene una distribución exponencial con parámetro λ. [*Sugerencia*: la función de distribución acumulativa de X es $F(x) = P(X \le x)$; $X \le x$ equivale a $U \le ?$]

b. ¿Cómo utilizaría el inciso (a) y un generador de números aleatorios para obtener valores observados derivados de una distribución exponencial con parámetro $\lambda = 10$?

124. Considere una variable aleatoria con media μ y desviación estándar σ, y sea $g(X)$ una función especificada de X. La aproximación de la serie de Taylor de primer orden a $g(X)$ en la cercanía de μ es

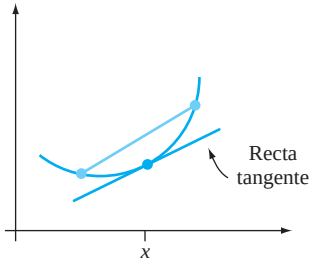
$$g(X) \approx g(\mu) + g'(\mu) \cdot (X - \mu)$$

El miembro del lado derecho de esta ecuación es una función lineal de X. Si la distribución de X está concentrada en un intervalo a lo largo del cual $g(\cdot)$ es aproximadamente lineal [p. ej., $\sqrt{x}$ es aproximadamente lineal en (1, 2)], entonces la ecuación produce aproximaciones a $E(g(X))$ y $V(g(X))$.

a. Dé expresiones para estas aproximaciones. [*Sugerencia*: use reglas de valor esperado y varianza para una función lineal $aX + b$.]

b. Si el voltaje v a través de un medio se mantiene fijo pero la corriente I es aleatoria, entonces la resistencia también será una variable aleatoria relacionada con I por $R = v/I$. Si $\mu_I = 20$ y $\sigma_I = .5$, calcule aproximaciones a μ_R y σ_R.

125. Una función $g(x)$ es *convexa* si la cuerda que conecta dos puntos cualesquiera de su gráfica quedan sobre ésta. Cuando $g(x)$ es diferenciable, una condición equivalente es que para cada x, la línea tangente en x queda por completo sobre o debajo de la gráfica. (Véase la figura a continuación.) ¿Cómo se compara $g(\mu) = g(E(X))$ con $E(g(X))$? [*Sugerencia*: la ecuación de la línea tangente en $x = \mu$ es $y = g(\mu) + g'(\mu) \cdot (x - \mu)$. Use la condición de convexidad, sustituya X por x y considere los valores esperados. [*Nota*: a menos que $g(x)$ sea lineal, la desigualdad resultante (por lo general llamada desigualdad de Jensen) es estricta ($<$ en lugar de $\le$); es válida tanto con variables aleatorias continuas como discretas.]



126. Si X tiene una distribución de Weibull con parámetros $\alpha = 2$ y β, demuestre que $Y = 2X^2/\beta^2$ tiene una distribución ji cuadrada con $\nu = 2$. [*Sugerencia*: la función de distribución acumulativa de Y es $P(Y \le y)$; exprese esta probabilidad en la forma $P(X \le g(y))$, use el hecho de que X tiene una función de distribución acumulativa de la forma de la expresión (4.12) y diferencie con respecto a y para obtener la función de densidad de probabilidad de Y.]

127. El registro crediticio de un individuo es un número calculado basado en el historial crediticio de dicho individuo el cual ayuda a un prestamista a determinar cuánto se le puede prestar o qué límite de crédito debe ser establecido para una tarjeta de crédito. Un artículo en *Los Ángeles Times* presentó datos que sugerían que una distribución beta con parámetros $A = 150$, $B = 850$, $\alpha = 8$, $\beta = 2$ proporcionaría una aproximación razonable a la distribución de registros de crédito estadounidenses. [*Nota*: los registros de crédito son valores enteros].

a. Sea X un registro estadounidense de crédito seleccionado al azar. ¿Cuáles son el valor medio y la desviación estándar de esta variable aleatoria? ¿Cuál es la probabilidad de que X esté dentro de 1 desviación estándar de su valor medio?

b. ¿Cuál es la probabilidad aproximada de que un registro seleccionado al azar excederá de 750 (lo que los prestamistas consideran un muy buen registro)?

128. Sea V el volumen de lluvia y W el volumen de escurrimiento (ambos en mm). De acuerdo con el artículo "Runoff Quality Analysis of Urban Catchments with Analytical Probability Models" (*J. of Water Resource Planning and Management*, 2006: 4–14), el volumen de escurrimiento será 0 si $V \le \nu_d$ y será $k(V - \nu_d)$ si $V > \nu_d$. Aquí ν_d es el volumen de almacenamiento en una depresión (una constante) y k (también una constante) es el coeficiente de escurrimiento. El artículo citado propone una distribución exponencial con parámetro λ para V.

a. Obtenga una expresión para la función de distribución acumulativa de W. [*Nota*: W no es puramente continua ni puramente discreta; en cambio tiene una distribución "combinada" con un componente discreto en 0 y es continua con valores $w > 0$.]

b. ¿Cuál es la función de densidad de probabilidad de W para $w > 0$? Úsela para obtener una expresión para el valor esperado de volumen de escurrimiento.

Bibliografía

Bury, Karl, *Statistical Distributions in Engineering*, Cambridge Univ. Press, Cambridge, Inglaterra, 1999. Un estudio informativo y fácil de leer de distribuciones y sus propiedades.

Johnson, Norman, Samuel Kotz y N. Balakrishnan, *Continuous Univariate Distributions*, vols. 1–2, Wiley, Nueva York, 1994. Estos dos volúmenes juntos presentan un estudio exhaustivo de varias distribuciones continuas.

Nelson, Wayne, *Applied Data Analysis*, Wiley, Nueva York, 1982. Presenta una amplia discusión de distribuciones y métodos que se utilizan en el análisis de datos de vida útil.

Olkin, Ingram, Cyrus Derman y Leon Gleser, *Probability Models and Applications* (2a. ed.), Macmillan, Nueva York, 1994. Una buena cobertura de las propiedades generales y distribuciones específicas.

5 Distribuciones de probabilidad conjunta y muestras aleatorias

INTRODUCCIÓN

En los capítulos 3 y 4 se estudiaron modelos de probabilidad para una sola variable aleatoria. Muchos problemas de probabilidad y estadística implican diversas variables aleatorias al mismo tiempo. En este capítulo, primero se discuten modelos de probabilidad del comportamiento conjunto (es decir, simultáneo) de diversas variables aleatorias, con énfasis especial en el caso en el cual las variables son independientes una de otra. En seguida se estudian los valores esperados de funciones de diversas variables aleatorias, incluidas la covarianza y la correlación como medidas del grado de asociación entre dos variables.

Las últimas tres secciones del capítulo consideran funciones de n variables aleatorias $X_1, X_2, \ldots, X_n$, con un enfoque especial en su promedio $(X_1 + \cdots + X_n)/n$. Tal función, por sí misma una variable aleatoria, recibe el nombre de *estadístico*. Se utilizan métodos de probabilidad para obtener información sobre la distribución de un estadístico. El resultado principal de este tipo es el teorema del límite central (TLC), la base de muchos procedimientos inferenciales que implican tamaños de muestra grandes.

5.1 Variables aleatorias conjuntamente distribuidas

Existen muchas situaciones experimentales en las cuales más de una variable aleatoria (rv) será de interés para un investigador. Primero se consideran las distribuciones de probabilidad conjunta para dos variables aleatorias discretas, en seguida para dos variables continuas y por último para más de dos variables.

Dos variables aleatorias discretas

La función de masa de probabilidad (fmp) de una sola variable aleatoria discreta X especifica cuánta masa de probabilidad está colocada en cada valor posible de X. La función de masa de probabilidad conjunta de dos variables aleatorias discretas X y Y describe cuánta masa de probabilidad se coloca en cada posible par de valores (x, y).

DEFINICIÓN

Sean X y Y dos variables aleatorias discretas definidas en el espacio muestral $\mathcal{S}$ de un experimento. La **función de masa de probabilidad conjunta** $p(x, y)$ se define para cada par de números (x, y) como

$$p(x, y) = P(X = x \text{ y } Y = y)$$

Debe ser el caso que $p(x, y) \geq 0$ y $\sum_x \sum_y p(x, y) = 1$.

Ahora sea A cualquier conjunto compuesto de pares de valores (x, y) (p. ej., $A = \{(x, y): x + y = 5\}$ o $\{(x, y): \text{máx}(x, y) \leq 3\}$). Entonces la probabilidad $P[(X, Y) \in A]$ se obtiene sumando la función de masa de probabilidad conjunta incluidos todos los pares en A:

$$P[(X, Y) \in A] = \sum_{(x, y)} \sum_{\in A} p(x, y)$$

Ejemplo 5.1 Una gran agencia de seguros presta servicios a numerosos clientes que han adquirido tanto una póliza de propietario de casa como una póliza de automóvil en la agencia. Por cada tipo de póliza, se debe especificar una cantidad deducible. Para una póliza de automóvil, las opciones son \$100 y \$250, mientras que para la póliza de propietario de casa, las opciones son 0, \$100 y \$200. Suponga que se selecciona al azar un individuo con ambos tipos de póliza de los archivos de la agencia. Sea X = la cantidad deducible sobre la póliza de auto y Y = la cantidad deducible sobre la póliza de propietario de casa. Los posibles pares (X, Y) son entonces (100, 0), (100, 100), (100, 200), (250, 0), (250, 100) y (250, 200); la función de masa de probabilidad conjunta especifica la probabilidad asociada con cada uno de estos pares, con cualquier otro par cuya probabilidad sea cero. Suponga que la **tabla de probabilidad conjunta** siguiente da la función de masa de probabilidad conjunta:

$p(x, y)$		y = 0	100	200
x	100	.20	.10	.20
	250	.05	.15	.30

Entonces $p(100, 100) = P(X = 100 \text{ y } Y = 100) = P(\$100 \text{ deducible sobre ambas pólizas}) = .10$. La probabilidad $P(Y \geq 100)$ se calcula sumando las probabilidades de todos los pares (x, y) para los cuales $y \geq 100$:

$$\begin{aligned} P(Y \geq 100) &= p(100, 100) + p(250, 100) + p(100, 200) + p(250, 200) \\ &= .75 \end{aligned}$$

■

Una vez que la función de masa de probabilidad conjunta de las dos variables X y Y está disponible en principio, es sencillo obtener la distribución de una sola de estas variables. Como ejemplo, sean X y Y el número de los cursos de estadística y de matemáticas, respectivamente, que se están cursando actualmente por una gran estadística seleccionada al azar. Supongamos que se quiere conocer la distribución de X y que cuando $X = 2$, los únicos valores posibles de Y son 0, 1 y 2. Entonces

$$\begin{aligned} p_X(2) = P(X = 2) &= P[(X, Y) = (2, 0) \text{ o } (2, 1) \text{ o } (2, 2)] \\ &= p(2, 0) + p(2, 1) + p(2, 2) \end{aligned}$$

Es decir, la función de masa de probabilidad conjunta se resume en todos los pares de la forma $(2, y)$. De manera más general, para cualquier posible valor x de X, la probabilidad $p_X(x)$ resulta de mantener x fija y sumar la función de masa de probabilidad conjunta $p(x, y)$ sobre toda y para la que el par (x, y) tiene una masa de probabilidad positiva. La misma estrategia se aplica a la obtención de la distribución de Y por sí misma.

DEFINICIÓN

La **función de masa de probabilidad marginal de *X*,** denotada por $p_X(x)$, está dada por

$$p_X(x) = \sum_{y:\, p(x,y)>0} p(x, y) \text{ para cada valor posible } x$$

De manera similar, la **función de masa de probabilidad marginal de *Y*** es

$$p_Y(y) = \sum_{x:\, p(x,y)>0} p(x, y) \text{ para cada valor posible } y.$$

El uso de la palabra *marginal* aquí es una consecuencia del hecho de que si la función de masa de probabilidad conjunta se muestra en una tabla rectangular como en el ejemplo 5.1, entonces los totales de los renglones dan como resultado la función de masa de probabilidad marginal de X y los totales de las columnas dan como resultado la función de masa de probabilidad marginal de Y. Una vez que estas fmp marginales están disponibles, se puede calcular la probabilidad de cualquier evento que involucre solamente a X o a Y.

Ejemplo 5.2
(Continuación del ejemplo 5.1)

Los valores posibles de X son $x = 100$ y $x = 250$, por lo que si se calculan los totales en los renglones de la tabla de probabilidad conjunta se obtiene

$$p_X(100) = p(100, 0) + p(100, 100) + p(100, 200) = .50$$

y

$$p_X(250) = p(250, 0) + p(250, 100) + p(250, 200) = .50$$

La función de masa de probabilidad marginal de X es entonces

$$p_X(x) = \begin{cases} .5 & x = 100, 250 \\ 0 & \text{de lo contrario} \end{cases}$$

Asimismo, la función de masa de probabilidad marginal de Y se obtiene con los totales de las columnas como

$$p_Y(y) = \begin{cases} .25 & y = 0, 100 \\ .50 & y = 200 \\ 0 & \text{de lo contrario} \end{cases}$$

Por lo tanto, $P(Y \geq 100) = p_Y(100) + p_Y(200) = .75$ como antes. ■

Dos variables aleatorias continuas

La probabilidad de que el valor observado de una variable aleatoria continua X esté en un conjunto unidimensional A (tal como un intervalo) se obtiene integrando la función de den-

sidad de probabilidad $f(x)$ a lo largo del conjunto A. Asimismo, la probabilidad de que el par (X, Y) de variables aleatorias continuas quede en un conjunto A en dos dimensiones (tal como un rectángulo) se obtiene integrando una función llamada *función de densidad conjunta*.

DEFINICIÓN

Sean X y Y variables aleatorias continuas. Una **función de densidad de probabilidad conjunta** $f(x, y)$ para estas dos variables es una función que satisface $f(x, y) \geq 0$ y $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\,dx\,dy = 1$. Entonces para cualquier conjunto A en dos dimensiones

$$P[(X, Y) \in A] = \iint_A f(x, y)\,dx\,dy$$

En particular, si A es el rectángulo bidimensional $\{(x, y): a \leq x \leq b, c \leq y \leq d\}$, entonces

$$P[(X, Y) \in A] = P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x, y)\,dy\,dx$$

Se puede considerar que $f(x, y)$ especifica una superficie situada a una altura $f(x, y)$ encima del punto (x, y) en un sistema de coordenadas tridimensional. Entonces $P[(X, Y) \in A]$ es el volumen debajo de esta superficie y sobre la región A, similar al área bajo una curva en el caso de una sola variable aleatoria. Esto se ilustra en la figura 5.1.

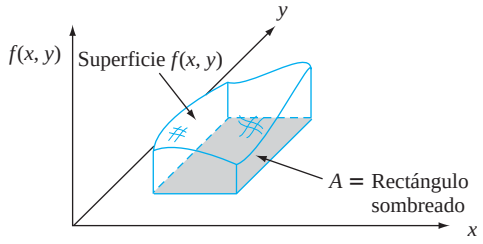


Figura 5.1 $P[(X, Y) \in A]$ = volumen bajo la superficie de densidad sobre A

Ejemplo 5.3 Un banco dispone tanto de una ventanilla para automovilistas como de una ventanilla normal. En un día seleccionado al azar, sea X = la proporción de tiempo que la ventanilla para automovilistas está en uso (por lo menos un cliente está siendo atendido o está esperando ser atendido) y Y = la proporción del tiempo que la ventanilla normal está en uso. Entonces el conjunto de valores posibles de (X, Y) es el rectángulo $D = \{(x, y): 0 \leq x \leq 1, 0 \leq y \leq 1\}$. Suponga que la función de densidad de probabilidad conjunta de (X, Y) está dada por

$$f(x, y) = \begin{cases} \frac{6}{5}(x + y^2) & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{de lo contrario} \end{cases}$$

Para verificar que ésta es una función de densidad de probabilidad legítima, obsérvese que $f(x, y) \geq 0$ y

$$\begin{aligned} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\,dx\,dy &= \int_0^1 \int_0^1 \frac{6}{5}(x + y^2)\,dx\,dy \\ &= \int_0^1 \int_0^1 \frac{6}{5}x\,dx\,dy + \int_0^1 \int_0^1 \frac{6}{5}y^2\,dx\,dy \\ &= \int_0^1 \frac{6}{5}x\,dx + \int_0^1 \frac{6}{5}y^2\,dy = \frac{6}{10} + \frac{6}{15} = 1 \end{aligned}$$

La probabilidad de que ninguna ventanilla esté ocupada más de un cuarto del tiempo es

$$P\left(0 \le X \le \frac{1}{4}, 0 \le Y \le \frac{1}{4}\right) = \int_0^{1/4}\int_0^{1/4} \frac{6}{5}(x + y^2)\,dx\,dy$$

$$= \frac{6}{5}\int_0^{1/4}\int_0^{1/4} x\,dx\,dy + \frac{6}{5}\int_0^{1/4}\int_0^{1/4} y^2\,dx\,dy$$

$$= \frac{6}{20}\cdot\frac{x^2}{2}\Big|_{x=0}^{x=1/4} + \frac{6}{20}\cdot\frac{y^3}{3}\Big|_{y=0}^{y=1/4} = \frac{7}{640}$$

$$= .0109$$

■

La función de densidad de probabilidad marginal de cada variable se puede obtener de una manera análoga a lo que se hizo en el caso de dos variables discretas. La función de densidad de probabilidad marginal de X en el valor x resulta de mantener x fija en el par (x, y) e *integrando* la función de densidad de probabilidad conjunta sobre y. La integración de la función de densidad de probabilidad conjunta con respecto a x da como resultado la función de densidad de probabilidad marginal de Y.

DEFINICIÓN

Las **funciones de densidad de probabilidad marginal** de X y Y, denotadas por $f_X(x)$ y $f_Y(y)$, respectivamente, están dadas por

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\,dy \quad \text{para } -\infty < x < \infty$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\,dx \quad \text{para } -\infty < y < \infty$$

Ejemplo 5.4 (Continuación del ejemplo 5.3)

La función de densidad de probabilidad marginal de X, la cual da la distribución de probabilidad del tiempo que permanece ocupada la ventanilla para automovilistas sin referencia a la ventanilla normal, es

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\,dy = \int_0^1 \frac{6}{5}(x + y^2)\,dy = \frac{6}{5}x + \frac{2}{5}$$

con $0 \le x \le 1$ y 0 de lo contrario. La función de densidad de probabilidad marginal de Y es

$$f_Y(y) = \begin{cases} \frac{6}{5}y^2 + \frac{3}{5} & 0 \le y \le 1 \\ 0 & \text{de lo contrario} \end{cases}$$

Entonces

$$P\left(\frac{1}{4} \le Y \le \frac{3}{4}\right) = \int_{1/4}^{3/4} f_Y(y)\,dy = \frac{37}{80} = .4625$$

■

En el ejemplo 5.3, la región de densidad conjunta positiva fue un rectángulo, lo que facilitó el cálculo de las funciones de densidad de probabilidad marginal. Considere ahora un ejemplo en el cual la región de densidad positiva es más complicada.

Ejemplo 5.5

Una compañía de nueces comercializa latas de nueces combinadas de lujo que contienen almendras, nueces de acajú y cacahuates. Suponga que el peso neto de cada lata es exactamente 1 lb, pero la contribución al peso de cada tipo de nuez es aleatoria. Como los tres pesos suman 1, un modelo de probabilidad conjunta de dos cualesquiera da toda la información necesaria sobre el peso del tercer tipo. Sea X = el peso de las almendras en una lata seleccionada y Y = el peso de las nueces de acajú. Entonces la región de densidad positiva es $D = \{(x, y): 0 \le x \le 1, 0 \le y \le 1, x + y \le 1\}$, la región sombreada ilustrada en la figura 5.2.

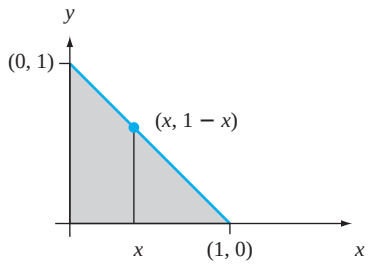


Figura 5.2 Región de densidad positiva para el ejemplo 5.5

Ahora sea la función de densidad de probabilidad conjunta de (X, Y)

$$f(x, y) = \begin{cases} 24xy & 0 \le x \le 1,\ 0 \le y \le 1,\ x + y \le 1 \\ 0 & \text{de lo contrario} \end{cases}$$

Con cualquier x fija, $f(x, y)$ se incrementa con y; con y fija, $f(x, y)$ se incrementa con x. Esto es apropiado porque las palabras *de lujo* implican que la mayor parte de lata deberá estar compuesta de almendras y nueces de acajú en lugar de cacahuates, así que la función de densidad deberá ser grande cerca del límite superior y pequeña cerca del origen. La superficie determinada por $f(x, y)$ se inclina hacia arriba desde cero a medida que (x, y) se alejan de uno u otro ejes.

Claramente, $f(x, y) \ge 0$. Para verificar la segunda condición sobre una función de densidad de probabilidad conjunta, recuérdese que una integral doble se calcula como una integral iterada manteniendo una variable fija (tal como x en la figura 5.2), integrando con los valores de la otra variable localizados a lo largo de la línea recta que pasa a través del valor de la variable fija y finalmente integrando todos los valores posibles de la variable fija. Así pues

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\,dy\,dx = \iint_D f(x, y)\,dy\,dx = \int_0^1 \left\{ \int_0^{1-x} 24xy\,dy \right\} dx$$

$$= \int_0^1 24x \left\{ \frac{y^2}{2}\Big|_{y=0}^{y=1-x} \right\} dx = \int_0^1 12x(1 - x)^2\,dx = 1$$

Para calcular la probabilidad de que los dos tipos de nueces conformen cuando mucho 50% de la lata, sea $A = \{(x, y): 0 \le x \le 1,\ 0 \le y \le 1, \text{ y } x + y \le .5\}$, como se muestra en la figura 5.3. Entonces

$$P((X, Y) \in A) = \iint_A f(x, y)\,dx\,dy = \int_0^{.5}\int_0^{.5-x} 24xy\,dy\,dx = .0625$$

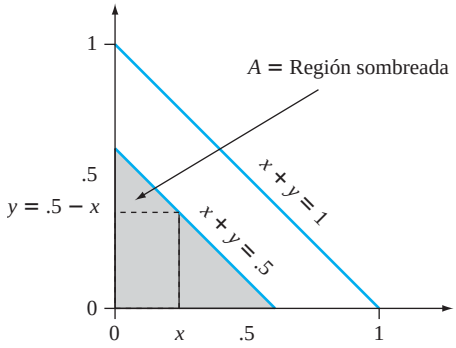


Figura 5.3 Cálculo de $P[(X, Y) \in A]$ para el ejemplo 5.5

La función de densidad de probabilidad marginal de las almendras se obtiene manteniendo X fija en x e integrando la función de densidad de probabilidad conjunta $f(x, y)$ a lo largo de la línea vertical que pasa por x:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \begin{cases} \displaystyle\int_0^{1-x} 24xy\, dy = 12x(1 - x)^2 & 0 \le x \le 1 \\ 0 & \text{de lo contrario} \end{cases}$$

Por simetría de $f(x, y)$ y la región D, la función de densidad de probabilidad marginal de Y se obtiene reemplazando x y X en $f_X(x)$ por y y Y, respectivamente. ■

Variables aleatorias independientes

En muchas situaciones, la información sobre el valor observado de una de las dos variables X y Y da información sobre el valor de la otra variable. En el ejemplo 5.1, la probabilidad marginal de X con $x = 250$ fue de .5, como lo fue la probabilidad de que $X = 100$. Si, no obstante, se dice que el individuo seleccionado tuvo $Y = 0$, entonces $X = 100$ es cuatro veces más probable que $X = 250$. Por lo tanto existe dependencia entre las dos variables.

En el capítulo 2 se señaló que una forma de definir la independencia de dos eventos es vía la condición de que $P(A \cap B) = P(A) \cdot P(B)$. A continuación se da una definición análoga de la independencia de dos variables aleatorias.

DEFINICIÓN

Se dice que dos variables aleatorias X y Y son **independientes** si por cada par de valores x y y

$$p(x, y) = p_X(x) \cdot p_Y(y) \qquad \text{cuando } X \text{ y } Y \text{ son discretas}$$

o (5.1)

$$f(x, y) = f_X(x) \cdot f_Y(y) \qquad \text{cuando } X \text{ y } Y \text{ son continuas}$$

Si (5.1) no se satisface con todos los pares (x, y), entonces se dice que X y Y son **dependientes.**

La definición dice que dos variables son independientes si su función de masa de probabilidad conjunta o función de densidad de probabilidad conjunta es el producto de las dos funciones de masa de probabilidad marginales o de las funciones de densidad de probabilidad marginales. Intuitivamente, la independencia dice que conocer el valor de una de las variables no proporciona información adicional acerca de cuál puede ser el valor de la otra variable.

Ejemplo 5.6 En la situación de la agencia de seguros de los ejemplos 5.1 y 5.2,

$$p(100, 100) = .10 \neq (.5)(.25) = p_X(100) \cdot p_Y(100)$$

de modo que X y Y no son independientes. La independencia de X y Y requiere que *toda* entrada en la tabla de probabilidad conjunta sea el producto de las probabilidades marginales que aparecen en los renglones y columnas correspondientes. ■

Ejemplo 5.7 (Continuación del ejemplo 5.5) Como $f(x, y)$ tiene la forma de un producto, X y Y parecerían ser independientes. Sin embargo, aunque $f_X\left(\frac{3}{4}\right) = f_Y\left(\frac{3}{4}\right) = \frac{9}{16}$, $f\left(\frac{3}{4}, \frac{3}{4}\right) = 0 \neq \frac{9}{16} \cdot \frac{9}{16}$ de modo que las variables no son en realidad independientes. Para que sean independientes $f(x, y)$ debe tener la forma $g(x) \cdot h(y)$ y la región de densidad positiva debe ser un rectángulo con sus lados paralelos a los ejes de coordenadas. ■

La independencia de dos variables aleatorias es más útil cuando la descripción del experimento en estudio sugiere que X y Y no tienen ningún efecto entre ellas. Entonces,

una vez que las funciones de masa de probabilidad o de densidad de probabilidad marginales han sido especificadas, la función de masa de probabilidad conjunta o la función de densidad de probabilidad conjunta es simplemente el producto de las dos funciones marginales. Se deduce que

$$P(a \le X \le b, c \le Y \le d) = P(a \le X \le b) \cdot P(c \le Y \le d)$$

Ejemplo 5.8 Suponga que las duraciones de dos componentes son independientes entre sí y que la distribución exponencial de la primera duración es X_1 con parámetro λ_1, mientras que la distribución exponencial de la segunda es X_2 con parámetro λ_2. Entonces la función de densidad de probabilidad conjunta es

$$\begin{aligned} f(x_1, x_2) &= f_{X_1}(x_1) \cdot f_{X_2}(x_2) \\ &= \begin{cases} \lambda_1 e^{-\lambda_1 x_1} \cdot \lambda_2 e^{-\lambda_2 x_2} = \lambda_1 \lambda_2 e^{-\lambda_1 x_1 - \lambda_2 x_2} & x_1 > 0, x_2 > 0 \\ 0 & \text{de lo contrario} \end{cases} \end{aligned}$$

Sean $\lambda_1 = 1/1000$ y $\lambda_2 = 1/1200$, de modo que las duraciones esperadas son 1000 y 1200 horas, respectivamente. La probabilidad de que ambas duraciones sean de por lo menos 1500 horas es

$$\begin{aligned} P(1500 \le X_1, 1500 \le X_2) &= P(1500 \le X_1) \cdot P(1500 \le X_2) \\ &= e^{-\lambda_1(1500)} \cdot e^{-\lambda_2(1500)} \\ &= (.2231)(.2865) = .0639 \end{aligned}$$

■

Más de dos variables aleatorias

Para modelar el comportamiento conjunto de más de dos variables aleatorias, se amplía el concepto de una distribución conjunta de dos variables.

DEFINICIÓN

Si $X_1, X_2, \ldots, X_n$ son variables aleatorias discretas, la función de masa de probabilidad conjunta de las variables es la función

$$p(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

Si las variables son continuas, la función de densidad de probabilidad conjunta de $X_1, \ldots, X_n$ es la función $f(x_1, x_2, \ldots, x_n)$ tal que para n intervalos cualesquiera $[a_1, b_1], \ldots, [a_n, b_n]$,

$$P(a_1 \le X_1 \le b_1, \ldots, a_n \le X_n \le b_n) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(x_1, \ldots, x_n)\, dx_n \ldots dx_1$$

En un experimento binomial, cada ensayo podría dar por resultado uno de sólo dos posibles resultados. Considérese ahora un experimento compuesto de n ensayos independientes e idénticos, en los que cada ensayo puede dar uno cualquiera de r posibles resultados. Sea $p_1 = P(\text{resultado } i \text{ en cualquier ensayo particular})$ y defínanse las variables aleatorias como $X_i =$ el número de ensayos que dan el resultado i ($i = 1, \ldots, r$). Tal experimento se llama **experimento multinomial** y la función de masa de probabilidad conjunta de $X_1, \ldots, X_r$ se llama **distribución multinomial.** Usando un argumento de conteo análogo al utilizado al deducir la distribución binomial, la función de masa de probabilidad conjunta de $X_1, \ldots, X_r$ se puede demostrar que es

$$\begin{aligned} &p(x_1, \ldots, x_r) \\ &= \begin{cases} \dfrac{n!}{(x_1!)(x_2!) \cdot \cdots \cdot (x_r!)} \; p_1^{x_1} \cdot \ldots \cdot p_r^{x_r} & x_i = 0, 1, 2, \ldots, \text{con } x_1 + \cdots + x_r = n \\ 0 & \text{de lo contrario} \end{cases} \end{aligned}$$

El caso $r = 2$ da como resultado la distribución binomial, con $X_1 =$ número de éxitos y $X_2 = n - X_1 =$ número de fallas.

Ejemplo 5.9 Si se determina el alelo de cada una de diez secciones de un chícharo obtenidas independientemente y $p_1 = P(\text{AA})$, $p_2 = P(\text{Aa})$, $p_3 = P(\text{aa})$, $X_1 =$ número de AA, $X_2 =$ número de Aa y $X_3 =$ número de aa, entonces la función de masa de probabilidad multinomial para estas X_i es

$$p(x_1, x_2, x_3) = \frac{10!}{(x_1!)(x_2!)(x_3!)} p_1^{x_1} p_2^{x_2} p_3^{x_3} \quad x_i = 0, 1, \ldots \text{ y } x_1 + x_2 + x_3 = 10$$

Con $p_1 = p_3 = .25$, $p_2 = .5$.

$$P(X_1 = 2, X_2 = 5, X_3 = 3) = p(2, 5, 3)$$
$$= \frac{10!}{2!\,5!\,3!}(.25)^2(.5)^5(.25)^3 = .0769$$ ■

Ejemplo 5.10 Cuando se utiliza cierto método para recolectar un volumen fijo de muestras de roca en una región, existen cuatro tipos de roca. Sean X_1, X_2 y X_3 la proporción por volumen de los tipos de roca 1, 2 y 3 en una muestra aleatoriamente seleccionada (la proporción del tipo de roca 4 es $1 - X_1 - X_2 - X_3$, de modo que una variable X_4 sería redundante). Si la función de densidad de probabilidad conjunta de X_1, X_2, X_3 es

$$f(x_1, x_2, x_3) = \begin{cases} kx_1x_2(1 - x_3) & 0 \le x_1 \le 1, 0 \le x_2 \le 1, 0 \le x_3 \le 1, x_1 + x_2 + x_3 \le 1 \\ 0 & \text{de lo contrario} \end{cases}$$

entonces k se determina como sigue

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1, x_2, x_3)\, dx_3\, dx_2\, dx_1$$
$$= \int_0^1 \left\{ \int_0^{1-x_1} \left[\int_0^{1-x_1-x_2} kx_1x_2(1 - x_3)\, dx_3 \right] dx_2 \right\} dx_1$$

El valor de la integral iterada es $k/144$, por lo tanto $k = 144$. La probabilidad de que las rocas de los tipos 1 y 2 integren más de 50% de la muestra es

$$P(X_1 + X_2 \le .5) = \int\int\int_{\substack{0 \le x_i \le 1 \text{ para } i=1,2,3 \\ x_1+x_2+x_3 \le 1,\ x_1+x_2 \le .5}} f(x_1, x_2, x_3)\, dx_3\, dx_2\, dx_1$$
$$= \int_0^{.5} \left\{ \int_0^{.5-x_1} \left[\int_0^{1-x_1-x_2} 144x_1x_2(1 - x_3)\, dx_3 \right] dx_2 \right\} dx_1$$
$$= .6066$$ ■

La noción de independencia de más de dos variables aleatorias es similar a la noción de independencia de más de dos eventos.

DEFINICIÓN

Se dice que las variables aleatorias $X_1, X_2, \ldots, X_n$ son **independientes** si para *cada* subconjunto $X_{i_1}, X_{i_2}, \ldots, X_{i_k}$ de las variables (cada par, cada terna, y así sucesivamente), la función de masa de probabilidad conjunta o la función de densidad de probabilidad conjunta del subconjunto es igual al producto de las funciones de masa de probabilidad o las funciones de densidad de probabilidad marginales.

Así pues, si las variables son independientes con $n = 4$, entonces la función de masa de probabilidad conjunta o la función de densidad de probabilidad conjunta de dos variables cualesquiera es el producto de las dos marginales, y asimismo para tres variables cualesquiera y cuatro variables juntas. Aún más importante, una vez que se dice que n variables son independientes, entonces la función de masa de probabilidad conjunta o la función de densidad de probabilidad conjunta es el producto de las n marginales.

Ejemplo 5.11 Si $X_1, \ldots, X_n$ representan las duraciones de n componentes y éstos operan de manera independiente uno de otro y cada duración está exponencialmente distribuida con parámetro λ, entonces para $x_1 \geq 0, x_2 \geq 0, \ldots, x_n \geq 0$,

$$f(x_1, x_2, \ldots, x_n) = (\lambda e^{-\lambda x_1}) \cdot (\lambda e^{-\lambda x_2}) \cdot \cdots \cdot (\lambda e^{-\lambda x_n}) = \lambda^n e^{-\lambda \Sigma x_i}$$

Si estos n componentes constituyen un sistema que fallará en cuanto un solo componente deje de funcionar, entonces la probabilidad de que el sistema dure más allá del tiempo t es

$$\begin{aligned} P(X_1 > t, \ldots, X_n > t) &= \int_t^\infty \cdots \int_t^\infty f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n \\ &= \left(\int_t^\infty \lambda e^{-\lambda x_1}\, dx_1 \right) \ldots \left(\int_t^\infty \lambda e^{-\lambda x_n}\, dx_n \right) \\ &= (e^{-\lambda t})^n = e^{-n\lambda t} \end{aligned}$$

Por consiguiente,

$$P(\text{duración del sistema } \leq t) = 1 - e^{-n\lambda t} \qquad \text{con } t \geq 0$$

lo que demuestra que la distribución de la duración del *sistema* es exponencial con parámetro $n\lambda$; el valor esperado de la duración del sistema es $1/n\lambda$. ■

En muchas situaciones experimentales que se considerarán en este libro, la independencia es una suposición razonable, de modo que la especificación de la distribución conjunta se reduce a decidir sobre distribuciones marginales apropiadas.

Distribuciones condicionales

Suponga $X =$ el número de defectos mayores en un automóvil nuevo seleccionado al azar y $Y =$ el número de defectos menores en el mismo auto. Si se sabe que el carro seleccionado tiene un defecto mayor, ¿cuál es ahora la probabilidad de que el carro tenga cuando mucho tres defectos menores?; es decir, ¿cuál es $P(Y \leq 3 \mid X = 1$? Asimismo, si X y Y denotan las duraciones de los neumáticos delantero y trasero de una motocicleta y sucede que $X = 10{,}000$ millas, ¿cuál es ahora la probabilidad de que Y sea cuando mucho de 15,000 millas y cuál es la duración esperada del neumático trasero "condicionada en" este valor de X? Preguntas de esta clase pueden ser respondidas estudiando distribuciones de probabilidad condicional.

DEFINICIÓN

Sean X y Y dos variables aleatorias continuas con función de densidad de probabilidad conjunta $f(x, y)$ y función de densidad de probabilidad marginal X, $f_X(x)$. Entonces para cualquier valor x de X para el cual $f_X(x) > 0$, la **función de densidad de probabilidad condicional de Y dado que $X = x$** es

$$f_{Y|X}(y \mid x) = \frac{f(x, y)}{f_X(x)} \qquad -\infty < y < \infty$$

Si X y Y son discretas, al reemplazar las funciones de densidad de probabilidad por funciones de masa de probabilidad en esta definición se obtiene la **función de masa de probabilidad condicional de Y cuando $X = x$.**

Obsérvese que la definición de $f_{Y \mid X}(y \mid x)$ es igual a la de $P(B \mid A)$, la probabilidad condicional de que B ocurra, dado que A ha ocurrido. Una vez que la función de densidad de probabilidad o la función de masa de probabilidad ha sido determinada, preguntas del tipo planteado al principio de esta subsección pueden ser respondidas integrando o sumando a lo largo de un conjunto apropiado de valores Y.

Ejemplo 5.12 Reconsidere la situación de los ejemplos 5.3 y 5.4 que implican X = la proporción del tiempo que la ventanilla para automovilista de un banco está ocupada y Y = la proporción análoga de la ventanilla normal. La función de densidad de probabilidad condicional de Y dado que $X = .8$ es

$$f_{Y|X}(y\,|\,.8) = \frac{f(.8, y)}{f_X(.8)} = \frac{1.2(.8 + y^2)}{1.2(.8) + .4} = \frac{1}{34}(24 + 30y^2) \quad 0 < y < 1$$

La probabilidad de que la ventanilla normal esté ocupada cuando mucho la mitad del tiempo dado que $X = .8$ es entonces

$$P(Y \le .5\,|\,X = .8) = \int_{-\infty}^{.5} f_{Y|X}(y\,|\,.8)\,dy = \int_{0}^{.5} \frac{1}{34}(24 + 30y^2)\,dy = .390$$

Utilizando la función de densidad de probabilidad marginal de Y se obtiene $P(Y \le .5) = .350$. Además $E(Y) = .6$, mientras que la proporción esperada del tiempo que la ventanilla normal está ocupada dado que $X = .8$ (una expectativa *condicional*) es

$$E(Y\,|\,X = .8) = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y\,|\,.8)\,dy = \frac{1}{34}\int_{0}^{1} y(24 + 30y^2)\,dy = .574$$ ■

Si las dos variables son independientes, las funciones de masa de probabilidad o de densidad de probabilidad marginales en el denominador, cancelarán el factor correspondiente en el numerador. La distribución condicional es entonces idéntica a la distribución marginal correspondiente.

EJERCICIOS Sección 5.1 (1–21)

1. Una gasolinera cuenta tanto con islas de autoservicio como de servicio completo. En cada isla hay una sola bomba de gasolina sin plomo regular con dos mangueras. Sea X el número de mangueras utilizadas en la isla de autoservicio en un tiempo particular y sea Y el número de mangueras en uso en la isla de servicio completo en ese tiempo. La función de masa de probabilidad conjunta de X y Y aparece en la tabla adjunta.

$p(x, y)$		y = 0	1	2
	0	.10	.04	.02
x	1	.08	.20	.06
	2	.06	.14	.30

a. ¿Cuál es $P(X = 1$ y $Y = 1)$?
b. Calcule $P(X \le 1$ y $Y \le 1)$.
c. Describa con palabras el evento $(X \ne 0$ y $Y \ne 0)$ y calcule su probabilidad.
d. Calcule la función de masa de probabilidad marginal de X y Y. Utilizando $p_X(x)$, ¿cuál es $P(X \le 1)$?
e. ¿Son X y Y variables aleatorias independientes? Explique.

2. Cuando un automóvil es detenido por una patrulla de seguridad, cada uno de los neumáticos es revisado en cuanto a desgaste y cada uno de los faros es revisado para ver si está apropiadamente alineado. Sean X el número de faros que necesitan ajuste y Y el número de neumáticos defectuosos.
a. Si X y Y son independientes con $p_X(0) = .5$, $p_X(1) = .3$, $p_X(2) = .2$ y $p_Y(0) = .6$, $p_Y(1) = .1$, $p_Y(2) = p_Y(3) = .05$ y $p_Y(4) = .2$, muestre la función de masa de probabilidad conjunta de (X, Y) en una tabla de probabilidad conjunta.
b. Calcule $P(X \le 1$ y $Y \le 1)$ con la tabla de probabilidad conjunta y verifique que es igual al producto $P(X \le 1) \cdot P(Y \le 1)$.
c. ¿Cuál es $P(X + Y = 0)$ (la probabilidad de ninguna violación)?
d. Calcule $P(X + Y \le 1)$.

3. Un supermercado cuenta tanto con una caja rápida como con una extrarrápida. Sea X_1 el número de clientes formados en la caja rápida a una hora particular del día y sea X_2 el número de clientes formados en la caja extrarrápida a la misma hora. Suponga que la función de masa de probabilidad conjunta de X_1 y X_2 es la que aparece en la tabla adjunta.

		x_2 = 0	1	2	3
	0	.08	.07	.04	.00
	1	.06	.15	.05	.04
x_1	2	.05	.04	.10	.06
	3	.00	.03	.04	.07
	4	.00	.01	.05	.06

a. ¿Cuál es $P(X_1 = 1, X_2 = 1)$; es decir, la probabilidad de que haya exactamente un cliente en cada caja?
b. ¿Cuál es $P(X_1 = X_2)$; es decir, la probabilidad de que los números de clientes en las dos cajas sean idénticos?
c. Sea A el evento en que hay por lo menos dos clientes más en una caja que en la otra. Exprese A en función de X_1 y X_2, y calcule la probabilidad de este evento.
d. ¿Cuál es la probabilidad de que el número total de clientes en las dos líneas sea exactamente cuatro? ¿Por lo menos cuatro?

4. Regrese a la situación descrita en el ejercicio 3.
- **a.** Determine la función de masa de probabilidad marginal de X_1 y en seguida calcule el número esperado de clientes formados en la caja rápida.
- **b.** Determine la función de masa de probabilidad marginal de X_2.
- **c.** Por inspección de las probabilidades $P(X_1 = 4)$, $P(X_2 = 0)$ y $P(X_1 = 4, X_2 = 0)$, ¿son X_1 y X_2 variables aleatorias independientes? Explique.

5. El número de clientes que esperan en el servicio de envoltura de regalos en una tienda de departamentos es una variable aleatoria X con valores posibles 0, 1, 2, 3, 4 y probabilidades correspondientes .1, .2, .3, .25, .15. Un cliente seleccionado al azar tendrá 1, 2 o 3 paquetes para envoltura con probabilidades de .6, .3 y .1, respectivamente. Sea Y = el número total de paquetes que van a ser envueltos para los clientes que esperan formados en la fila (suponga que el número de paquetes entregado por un cliente es independiente del número entregado por cualquier otro cliente).
- **a.** Determine $P(X = 3, Y = 3)$, es decir, $p(3, 3)$.
- **b.** Determine $p(4, 11)$.

6. Sea X el número de cámaras digitales Canon vendidas durante una semana particular por una tienda. La función de masa de probabilidad de X es

x	0	1	2	3	4
$p_X(x)$	.1	.2	.3	.25	.15

60% de todos los clientes que compran estas cámaras también compran una garantía extendida. Sea Y el número de compradores durante esta semana que compran una garantía extendida.
- **a.** ¿Cuál es $P(X = 4, Y = 2)$? [*Sugerencia*: esta probabilidad es igual a $P(Y = 2 \mid X = 4) \cdot P(X = 4)$; ahora piense en las cuatro compras como cuatro ensayos de un experimento binomial, con el éxito en un ensayo correspondiente a comprar una garantía extendida.]
- **b.** Calcule $P(X = Y)$.
- **c.** Determine la función de masa de probabilidad conjunta de X y Y y luego la función de masa de probabilidad marginal de Y.

7. La distribución de probabilidad conjunta del número X de carros y el número Y de autobuses por ciclo de señal en un carril de vuelta a la izquierda propuesto se muestra en la tabla de probabilidad conjunta anexa.

$p(x, y)$		y = 0	1	2
	0	.025	.015	.010
	1	.050	.030	.020
	2	.125	.075	.050
x	3	.150	.090	.060
	4	.100	.060	.040
	5	.050	.030	.020

- **a.** ¿Cuál es la probabilidad de que haya exactamente un carro y exactamente un autobús durante un ciclo?
- **b.** ¿Cuál es la probabilidad de que haya cuando mucho un carro y cuando mucho un autobús durante un ciclo?
- **c.** ¿Cuál es la probabilidad de que haya exactamente un carro durante un ciclo? ¿Exactamente un autobús?
- **d.** Suponga que el carril de vuelta a la izquierda tiene una capacidad de cinco carros y que un autobús equivale a tres carros. ¿Cuál es la probabilidad de un exceso de vehículos durante un ciclo?
- **e.** ¿Son X y Y variables aleatorias independientes? Explique.

8. Un almacén cuenta con 30 componentes de un tipo, de los cuales 8 fueron surtidos por el proveedor 1, 10 por el proveedor 2 y 12 por el proveedor 3. Seis de éstos tienen que ser seleccionados al azar para un ensamble particular. Sea X = el número de componentes del proveedor 1 seleccionados, Y = el número de componentes del proveedor 2 seleccionados y que $p(x, y)$ denote la función de masa de probabilidad conjunta de X y Y.
- **a.** ¿Cuál es $p(3, 2)$? [*Sugerencia*: la probabilidad de que cada muestra de tamaño 6 sea seleccionada es igual. Por consiguiente, $p(3, 2)$ = (número de resultados con $X = 3$ y $Y = 2$)/(el número total de resultados). Ahora use la regla de producto para el conteo para obtener el numerador y el denominador.]
- **b.** Utilizando la lógica del inciso (a), obtenga $p(x, y)$. (Esto puede ser considerado como un muestreo con distribución hipergeométrica multivariante sin reemplazo de una población finita compuesta por más de dos categorías.)

9. Se supone que cada neumático delantero de un tipo particular de vehículo está inflado a una presión de 26 lb/pulg^2. Suponga que la presión de aire real en cada neumático es una variable aleatoria, X para el neumático derecho y Y para el izquierdo con función de densidad de probabilidad conjunta

$$f(x, y) = \begin{cases} K(x^2 + y^2) & 20 \le x \le 30,\ 20 \le y \le 30 \\ 0 & \text{de lo contrario} \end{cases}$$

- **a.** ¿Cuál es el valor de K?
- **b.** ¿Cuál es la probabilidad de que ambos neumáticos estén inflados a menos presión?
- **c.** ¿Cuál es la probabilidad de que la diferencia en la presión del aire entre los dos neumáticos sea de cuando mucho 2 lb/pulg^2?
- **d.** Determine la distribución (marginal) de la presión del aire sólo en el neumático derecho.
- **e.** ¿Son X y Y variables aleatorias independientes?

10. Annie y Alvie acordaron encontrarse entre las 5:00 p.m. y las 6:00 p.m. para cenar en un restaurante local de comida saludable. Sea X = la hora de llegada de Annie y Y = la hora de llegada de Alvie. Suponga que X y Y son independientes, cada una distribuida uniformemente en el intervalo [5, 6].
- **a.** ¿Cuál es la función de densidad de probabilidad conjunta de X y Y?
- **b.** ¿Cuál es la probabilidad de que ambas lleguen entre las 5:15 y las 5:45?
- **c.** Si la primera en llegar espera sólo 10 min antes de irse a comer a otra parte, ¿cuál es la probabilidad de que cenen en el restaurante de comida saludable? [*Sugerencia*: el evento de interés es $A = \{(x, y): |x - y| \le \frac{1}{6}\}$.]

11. Dos profesores acaban de entregar los exámenes finales para su copia. Sea X el número de errores tipográficos en el examen del primer profesor y Y el número de tales errores en el segundo examen. Suponga que X tiene una distribución de Poisson con parámetro μ_1, que Y tiene una distribución de Poisson con parámetro μ_2 y que X y Y son independientes.

a. ¿Cuál es la función de masa de probabilidad conjunta de X y Y?
b. ¿Cuál es la probabilidad de que se cometa cuando mucho un error en ambos exámenes combinados?
c. Obtenga una expresión general para la probabilidad de que el número total de errores en los dos exámenes sea m (donde m es un entero no negativo). [*Sugerencia*: $A = \{(x, y): x + y = m\} = \{(m, 0), (m - 1, 1), \ldots, (1, m - 1), (0, m)\}$. Ahora sume la función de masa de probabilidad conjunta a lo largo de $(x, y) \in A$ y use el teorema binomial, el cual dice que

$$\sum_{k=0}^{m} \binom{m}{k} a^k b^{m-k} = (a + b)^m$$

para cualquier a, b.]

12. Dos componentes de una minicomputadora tienen la siguiente función de densidad de probabilidad conjunta de sus vidas útiles X y Y:

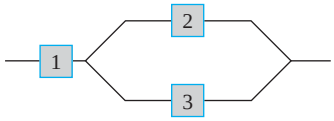
$$f(x, y) = \begin{cases} xe^{-x(1+y)} & x \geq 0 \text{ y } y \geq 0 \\ 0 & \text{de lo contrario} \end{cases}$$

a. ¿Cuál es la probabilidad de que la vida útil X del primer componente exceda de 3?
b. ¿Cuáles son las funciones de densidad de probabilidad marginales de X y Y? ¿Son las dos vidas útiles independientes? Explique.
c. ¿Cuál es la probabilidad de que la vida útil de por lo menos un componente exceda de 3?

13. Tiene dos focos para una lámpara particular. Sea X = la vida útil del primer foco y Y = la vida útil del segundo (ambas en miles de horas). Suponga que X y Y son independientes y que cada una tiene una distribución exponencial con parámetro $\lambda = 1$.
a. ¿Cuál es la función de densidad de probabilidad conjunta de X y Y?
b. ¿Cuál es la probabilidad de que cada foco dure cuando mucho 1000 horas (es decir, $X \leq 1$ y $Y \leq 1$)?
c. ¿Cuál es la probabilidad de que la vida útil total de los dos focos sea cuando mucho de 2? [*Sugerencia*: trace una figura de la región $A = \{(x, y): x \geq 0, y \geq 0, x + y \leq 2\}$ antes de integrar.]
d. ¿Cuál es la probabilidad de que la vida útil total sea de entre 1 y 2?

14. Suponga que tiene diez focos y que la vida útil de cada uno es independiente de la de los demás y que cada vida útil tiene una distribución exponencial con parámetro λ.
a. ¿Cuál es la probabilidad de que los diez focos fallen antes del tiempo t?
b. ¿Cuál es la probabilidad de que exactamente k de los diez focos fallen antes del tiempo t?
c. Suponga que nueve de los focos tienen vidas útiles exponencialmente distribuidas con parámetro λ y que el foco restante tiene una vida útil que está exponencialmente distribuida con parámetro θ (fue hecho por otro fabricante). ¿Cuál es la probabilidad de que exactamente cinco de los diez focos fallen antes del tiempo t?

15. Considere un sistema compuesto de tres componentes como se ilustra. El sistema continuará funcionando en tanto el primer componente funcione y el componente 2 o el componente 3 funcionen. Sean X_1, X_2 y X_3 las vidas útiles de los componentes 1, 2 y 3, respectivamente. Suponga que las X_i son independientes una de otra y que cada X_i tiene una distribución exponencial con parámetro λ.



a. Sea Y la vida útil del sistema. Obtenga la función de distribución acumulativa de Y y diferénciela para obtener la función de densidad de probabilidad. [*Sugerencia*: $F(y) = P(Y \leq y)$; exprese el evento $\{Y \leq y\}$ en función de uniones y/o intersecciones de los tres eventos $\{X_1 \leq y\}$, $\{X_2 \leq y\}$ y $\{X_3 \leq y\}$.]
b. Calcule la vida útil esperada del sistema.

16. **a.** Para $f(x_1, x_2, x_3)$ del ejemplo 5.10, calcule la **función de densidad marginal conjunta** sólo de X_1 y X_3 (integrando para x_2).
b. ¿Cuál es la probabilidad de que rocas de los tipos 1 y 3 constituyan cuando mucho 50% de la muestra? [*Sugerencia*: use el resultado del inciso (a).]
c. Calcule la función de densidad de probabilidad marginal sólo de X_1 [*Sugerencia*: use el resultado del inciso (a).]

17. Un ecólogo desea seleccionar un punto dentro de una región de muestreo circular de acuerdo con una distribución uniforme (en la práctica esto podría hacerse seleccionando primero una dirección y luego una distancia a partir del centro en esa dirección). Sea X = la coordenada x del punto seleccionado y Y = la coordenada y del punto seleccionado. Si el círculo tiene su centro en (0, 0) y su radio es R, entonces la función de densidad de probabilidad conjunta de X y Y es

$$f(x, y) = \begin{cases} \dfrac{1}{\pi R^2} & x^2 + y^2 \leq R^2 \\ 0 & \text{de lo contrario} \end{cases}$$

a. ¿Cuál es la probabilidad de que el punto seleccionado quede dentro de $R/2$ del centro de la región circular? [*Sugerencia*: trace una figura de la región de densidad positiva D. Como $f(x, y)$ es constante en D, el cálculo de probabilidad se reduce al cálculo de un área.]
b. ¿Cuál es la probabilidad de que tanto X como Y difieran de 0 por cuando mucho $R/2$?
c. Responda el inciso (b) con $R/\sqrt{2}$ reemplazando a $R/2$.
d. ¿Cuál es la función de densidad de probabilidad marginal de X? ¿De Y? ¿Son X y Y independientes?

18. Remítase al ejercicio 1 y responda las siguientes preguntas:
a. Dado que $X = 1$, determine la función de masa de probabilidad condicional de Y; es decir, $p_{Y \mid X}(0 \mid 1)$, $p_{Y \mid X}(1 \mid 1)$ y $p_{Y \mid X}(2 \mid 1)$.
b. Dado que dos mangueras están en uso en la isla de autoservicio, ¿cuál es la función de masa de probabilidad condicional del número de mangueras en uso en la isla de servicio completo?
c. Use el resultado del inciso (b) para calcular la probabilidad condicional $P(Y \leq 1 \mid X = 2)$.
d. Dado que dos mangueras están en uso en la isla de servicio completo, ¿cuál es la función de masa de probabilidad condicional del número en uso en la isla de autoservicio?

19. La función de densidad de probabilidad conjunta de las presiones de los neumáticos delanteros derecho e izquierdo se da en el ejercicio 9.

a. Determine la función de densidad de probabilidad condicional de Y dado que $X = x$ y la función de densidad de probabilidad condicional de X dado que $Y = y$.

b. Si la presión del neumático derecho es de 22 lb/pulg^2, ¿cuál es la probabilidad de que la presión del neumático izquierdo sea de por lo menos 25 lb/pulg^2? Compare esto con $P(Y \geq 25)$.

c. Si la presión del neumático derecho es de 22 lb/pulg^2, ¿cuál es la presión esperada en el neumático izquierdo y cuál es la desviación estándar de la presión en este neumático?

20. Sean X_1, X_2, X_3, X_4, X_5 y X_6 los números de caramelos M&M azules, cafés, verdes, naranjas, rojos y amarillos, respectivamente, en una muestra de tamaño n. Entonces estas X_i tienen una distribución multinomial. De acuerdo con el sitio web de M&M, las proporciones de colores son $p_1 = .24$, $p_2 = .13$, $p_3 = .16$, $p_4 = .20$, $p_5 = .13$ y $p_6 = .14$.

a. Si $n = 12$, ¿cuál es la probabilidad de que haya exactamente dos caramelos M&M de cada color?

b. Con $n = 20$, ¿cuál es la probabilidad de que haya cuando mucho cinco caramelos naranja? [*Sugerencia*: considere el caramelo naranja como un éxito y cualquier otro color como falla.]

c. En una muestra de 20 caramelos M&M, ¿cuál es la probabilidad de que el número de caramelos azules, verdes o naranjas sea por lo menos de 10?

21. Sean X_1, X_2 y X_3 las vidas útiles de los componentes 1, 2 y 3 en un sistema de tres componentes.

a. ¿Cómo definiría la función de densidad de probabilidad condicional de X_3 dado que $X_1 = x_1$ y $X_2 = x_2$?

b. ¿Cómo definiría la función de densidad de probabilidad conjunta condicional de X_2 y X_3 dado que $X_1 = x_1$?

5.2 Valores esperados, covarianza y correlación

Previamente se vio que cualquier función $h(x)$ de una sola variable aleatoria X es por sí misma una variable aleatoria. Sin embargo, para calcular $E[h(X)]$, no es necesario obtener la distribución de probabilidad de $h(X)$; en cambio, $E[h(X)]$ se calculó como un promedio ponderado de valores de $h(x)$, donde la función de ponderación fue la función de masa de probabilidad $p(x)$ o la función de densidad de probabilidad $f(x)$ de X. Se obtiene un resultado similar para una función $h(X, Y)$ de dos variables aleatorias conjuntamente distribuidas.

PROPOSICIÓN

Sean X y Y variables aleatorias conjuntamente distribuidas con función de masa de probabilidad $p(x, y)$ o función de densidad de probabilidad $f(x, y)$ ya sea que las variables sean discretas o continuas. Entonces el valor esperado de una función $h(X, Y)$ denotada por $E[h(X, Y)]$ o $\mu_{h(X, Y)}$ está dada por

$$E[h(X, Y)] = \begin{cases} \sum_x \sum_y h(x, y) \cdot p(x, y) & \text{si } X \text{ y } Y \text{ son discretas} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) \cdot f(x, y)\, dx\, dy & \text{si } X \text{ y } Y \text{ son continuas} \end{cases}$$

Ejemplo 5.13 Cinco amigos compraron boletos para un concierto. Si los boletos son para los asientos 1–5 en una fila particular y los boletos se distribuyen al azar entre los cinco, ¿cuál es el número esperado de asientos que separen a cualesquiera dos de los cinco? Sean X y Y los números de asiento del primero y segundo individuos, respectivamente. Los pares posibles (X, Y) son $\{(1, 2), (1, 3), \ldots, (5, 4)\}$ y la función de masa de probabilidad conjunta de (X, Y) es

$$p(x, y) = \begin{cases} \frac{1}{20} & x = 1, \ldots, 5; y = 1, \ldots, 5; x \neq y \\ 0 & \text{de lo contrario} \end{cases}$$

El número de asientos que separan a los dos individuos es $h(X, Y) = |X - Y| - 1$. La tabla adjunta da $h(x, y)$ para cada posible par (x, y).

$h(x, y)$		x 1	2	3	4	5
	1	—	0	1	2	3
	2	0	—	0	1	2
y	3	1	0	—	0	1
	4	2	1	0	—	0
	5	3	2	1	0	—

Por lo tanto

$$E[h(X, Y)] = \sum_{(x, y)} \sum h(x, y) \cdot p(x, y) = \sum_{\substack{x=1 \\ x \neq y}}^{5} \sum_{y=1}^{5} (|x - y| - 1) \cdot \frac{1}{20} = 1$$

Ejemplo 5.14 En el ejemplo 5.5, la función de densidad de probabilidad conjunta de la cantidad X de almendras y la cantidad Y de nueces de acajú en una lata de 1 lb de nueces fue

$$f(x, y) = \begin{cases} 24xy & 0 \leq x \leq 1, 0 \leq y \leq 1, x + y \leq 1 \\ 0 & \text{de lo contrario} \end{cases}$$

Si 1 lb de almendras le cuesta a la compañía \$1.00, 1 lb de nuez de acajú le cuesta \$1.50 y 1 lb de cacahuates le cuesta \$.50, entonces el costo total del contenido de una lata es

$$h(X, Y) = (1)X + (1.5)Y + (.5)(1 - X - Y) = .5 + .5X + Y$$

(puesto que $1 - X - Y$ del peso se compone de cacahuates). El costo esperado total es

$$\begin{aligned} E[h(X, Y)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) \cdot f(x, y) dx\, dy \\ &= \int_{0}^{1} \int_{0}^{1-x} (.5 + .5x + y) \cdot 24xy\, dy\, dx = \$1.10 \end{aligned}$$

El método de calcular el valor esperado de una función $h(X_1, \ldots, X_n)$ de n variables aleatorias es similar al de dos variables aleatorias. Si las X_i son discretas, $E[h(X_1, \ldots, X_n)]$ es una suma de n dimensiones; si las X_i son continuas, es una integral de n dimensiones.

Covarianza

Cuando dos variables aleatorias X y Y no son independientes, con frecuencia es de interés valorar qué tan fuerte están relacionadas una con otra.

DEFINICIÓN

La **covarianza** entre dos variables aleatorias X y Y es

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

$$= \begin{cases} \sum_x \sum_y (x - \mu_X)(y - \mu_Y) p(x, y) & X, Y \text{ discretas} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f(x, y)\, dx\, dy & X, Y \text{ continuas} \end{cases}$$

Es decir, como $X - \mu_X$ y $Y - \mu_Y$ son las desviaciones de las dos variables con respecto a sus valores medios, la covarianza es el producto esperado de las desviaciones. Obsérvese que $\text{Cov}(X, X) = E[(X - \mu_X)^2] = V(X)$.

La exposición razonada para la definición es como sigue. Suponga que X y Y tienen una fuerte relación positiva entre ellas, lo que significa que los valores grandes de X tienden a ocurrir con valores grandes de Y y los valores pequeños de X con los valores pequeños de Y. Entonces la mayor parte de la masa o densidad de probabilidad estará asociada para $(x - \mu_X)$ y $(y - \mu_Y)$, ambos positivos (tanto X como Y por arriba de sus respectivas medias) o ambos negativos, así que el producto $(x - \mu_X)(y - \mu_Y)$ tenderá a ser positivo. Por tanto para una fuerte relación positiva, $\text{Cov}(X, Y)$ deberá ser bastante positiva. Con una fuerte relación negativa los signos de $(x - \mu_X)$ y $(y - \mu_Y)$ tenderán a ser opuestos, lo que da un producto negativo. Por tanto, con una fuerte relación negativa, $\text{Cov}(X, Y)$ deberá ser bastante negativa. Si X y Y no están fuertemente relacionadas, los productos positivo y negativo tenderán a eliminarse entre sí, lo que da una covarianza de cerca de 0. La figura 5.4 ilustra las diferentes posibilidades. La covarianza depende *tanto* del conjunto de pares posibles *como* de las probabilidades. En la figura 5.4, las probabilidades podrían ser cambiadas sin que se altere el conjunto de pares posibles y esto podría cambiar drásticamente el valor de $\text{Cov}(X, Y)$.

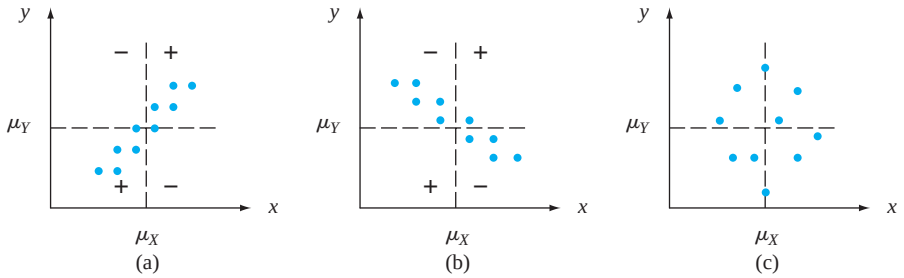


Figura 5.4 $p(x, y) = 1/10$ de cada uno de los diez pares correspondientes a los puntos indicados; (a) covarianza positiva; (b) covarianza negativa; (c) covarianza cerca de cero

Ejemplo 5.15 Las funciones de masa de probabilidad conjunta y marginal de X = cantidad deducible sobre una póliza de automóvil y Y = cantidad deducible sobre una póliza de propietario de casa en el ejemplo 5.1 fueron

$p(x, y)$		y = 0	100	200
x	100	.20	.10	.20
	250	.05	.15	.30

x	100	250
$p_X(x)$	.5	.5

y	0	100	200
$p_Y(y)$	.25	.25	.5

de las cuales $\mu_X = \Sigma x p_X(x) = 175$ y $\mu_Y = 125$. Por consiguiente,

$$\text{Cov}(X, Y) = \sum_{(x, y)} \sum (x - 175)(y - 125)p(x, y)$$

$$= (100 - 175)(0 - 125)(.20) + \cdots + (250 - 175)(200 - 125)(.30)$$

$$= 1875$$

■

La siguiente fórmula abreviada para $\text{Cov}(X, Y)$ simplifica los cálculos.

PROPOSICIÓN

$$\text{Cov}(X, Y) = E(XY) - \mu_X \cdot \mu_Y$$

De acuerdo con esta fórmula, no se requieren sustracciones intermedias; sólo al final del cálculo se resta $\mu_X \cdot \mu_Y$ de $E(XY)$. La comprobación implica expandir $(X - \mu_X)(Y - \mu_Y)$ y luego considerar el valor esperado de cada término por separado.

Ejemplo 5.16 (Continuación del ejemplo 5.5)

Las funciones de densidad de probabilidad conjunta y marginal de X = cantidad de almendras y Y = cantidad de nueces de acajú fueron

$$f(x, y) = \begin{cases} 24xy & 0 \le x \le 1, 0 \le y \le 1, x + y \le 1 \\ 0 & \text{de lo contrario} \end{cases}$$

$$f_X(x) = \begin{cases} 12x(1 - x)^2 & 0 \le x \le 1 \\ 0 & \text{de lo contrario} \end{cases}$$

con $f_Y(y)$ obtenida reemplazando x por y en $f_X(x)$. Es fácil verificar que $\mu_X = \mu_Y = \frac{2}{5}$ y

$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy\, f(x, y)\, dx\, dy = \int_0^1 \int_0^{1-x} xy \cdot 24xy\, dy\, dx$$

$$= 8 \int_0^1 x^2(1 - x)^3\, dx = \frac{2}{15}$$

Por lo tanto $\text{Cov}(X, Y) = \frac{2}{15} - \left(\frac{2}{5}\right)\left(\frac{2}{5}\right) = \frac{2}{15} - \frac{4}{25} = -\frac{2}{75}$. Una covarianza negativa se considera razonable en este caso porque más almendras contenidas en la lata implican menos nueces de acajú. ■

Pudiera parecer que la relación en el ejemplo de los seguros es bastante fuerte puesto que $\text{Cov}(X, Y) = 1875$, mientras que $\text{Cov}(X, Y) = -\frac{2}{75}$ en el ejemplo de las nueces parecería implicar una relación bastante débil. Desafortunadamente, la covarianza tiene un serio defecto que hace imposible interpretar un valor calculado. En el ejemplo de los seguros, suponga que la cantidad deducible se expresó en centavos en lugar de en dólares. Entonces $100X$ reemplazaría a X, $100Y$ reemplazaría a Y y la covarianza resultante sería $\text{Cov}(100X, 100Y) = (100)(100)\text{Cov}(X, Y) = 18{,}750{,}000$. Si, por otra parte, la cantidad deducible se hubiera expresado en cientos de dólares, la covarianza calculada habría sido $(.01)(.01)(1875) = .1875$. *El defecto de la covarianza es que su valor calculado depende críticamente de las unidades de medición*. De manera ideal, la selección de las unidades no debe tener efecto en la medida de la fuerza de la relación. Esto se logra graduando a escala la covarianza.

Correlación

DEFINICIÓN

El **coeficiente de correlación** de X y Y, denotado por $\text{Corr}(X, Y)$, $\rho_{X,Y}$, o simplemente ρ, está definido por

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

Ejemplo 5.17

Es fácil verificar que en el escenario de los seguros del ejemplo 5.15, $E(X^2) = 36{,}250$, $\sigma_X^2 = 36{,}250 - (175)^2 = 5625$, $\sigma_X = 75$, $E(Y^2) = 22{,}500$, $\sigma_Y^2 = 6875$ y $\sigma_Y = 82.92$. Esto da

$$\rho = \frac{1875}{(75)(82.92)} = .301$$

■

La siguiente proposición muestra que ρ remedia el defecto de Cov(X, Y) y también sugiere cómo reconocer la existencia de una fuerte relación (lineal).

PROPOSICIÓN

1. Si a y c son ambas positivas o ambas negativas,

$$\text{Corr}(aX + b, cY + d) = \text{Corr}(X, Y)$$

2. Para dos variables aleatorias X y Y cualesquiera, $-1 \leq \text{Corr}(X, Y) \leq 1$.

La proposición 1 dice precisamente que el coeficiente de correlación no se ve afectado por un cambio lineal en las unidades de medición (si, por ejemplo, X = temperatura en °C, entonces $9X/5 + 32$ = temperatura en °F). De acuerdo con la proposición 2, la relación positiva más fuerte posible es puesta en evidencia por $\rho = +1$, en tanto que la relación negativa más fuerte posible corresponde a $\rho = -1$. La comprobación de la primera proposición se ilustra en el ejercicio 35, y la de la segunda aparece en el ejercicio suplementario 87 al final del capítulo. Para propósitos descriptivos, la relación se describirá como fuerte si $|\rho| \geq .8$, moderada si $.5 < |\rho| < .8$ y débil si $|\rho| \leq .5$.

Si se considera que $p(x, y)$ o $f(x, y)$ prescribe un modelo matemático de cómo las dos variables numéricas X y Y están distribuidas en alguna población (estatura y peso, calificaciones del examen de aptitud escolar verbal y cuantitativa, etc.), entonces ρ es una característica o parámetro de población que mide cuán fuertemente están relacionadas X y Y en la población. En el capítulo 12 se considerará tomar una muestra de pares $(x_1, y_1), \ldots, (x_n, y_n)$ de la población. El coeficiente de correlación muestral r se definirá y utilizará entonces para hacer inferencias con respecto a ρ.

El coeficiente de correlación ρ no es en realidad una medida completamente general de la fuerza de una relación.

PROPOSICIÓN

1. Si X y Y son independientes, entonces $\rho = 0$, pero $\rho = 0$ no implica independencia.

2. $\rho = 1$ o -1 si y sólo si $Y = aX + b$ con algunos números a y b con $a \neq 0$.

Esta proposición dice que ρ mide el grado de relación **lineal** entre X y Y, y sólo cuando las dos variables están perfectamente relacionadas de una manera lineal, ρ será tan positivo o negativo como pueda serlo. Un ρ menor que 1 en valor absoluto indica sólo que la relación no es completamente lineal, pero que aún puede haber una fuerte relación no lineal. Además, $\rho = 0$ no implica que X y Y sean independientes, sino sólo que existe una ausencia completa de relación lineal. Cuando $\rho = 0$, se dice que X y Y **no están correlacionadas.** Dos variables podrían no estar correlacionadas y no obstante ser altamente dependientes porque existe una fuerte relación no lineal, así que se debe tener cuidado de no concluir demasiado del hecho de que $\rho = 0$.

Ejemplo 5.18 Sean X y Y variables aleatorias discretas con función de masa de probabilidad conjunta

$$p(x, y) = \begin{cases} \frac{1}{4} & (x, y) = (-4, 1), (4, -1), (2, 2), (-2, -2) \\ 0 & \text{de lo contrario} \end{cases}$$

Los puntos que reciben masa de probabilidad positiva están identificados en el sistema de coordenadas (x, y) en la figura 5.5. Es evidente por la figura que el valor de X está completamente determinado por el valor de Y y viceversa, de modo que las dos variables son completamente dependientes. Sin embargo, por simetría $\mu_X = \mu_Y = 0$ y $E(XY) = (-4)\frac{1}{4} + (-4)\frac{1}{4} + (4)\frac{1}{4} + (4)\frac{1}{4} = 0$, la covarianza es por tanto $\text{Cov}(X,Y) = E(XY) - \mu_X \cdot \mu_Y = 0$ y por consiguiente $\rho_{X,Y} = 0$. ¡Aunque hay una dependencia perfecta, también hay una ausencia completa de cualquier relación lineal!

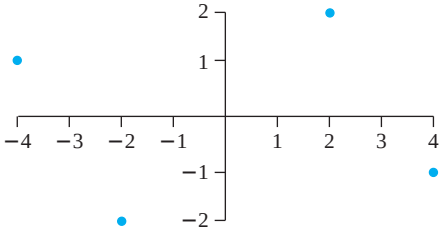


Figura 5.5 Población de pares del ejemplo 5.18 ■

Un valor de ρ próximo a 1 no necesariamente implica que el incremento del valor de X *hace* que se incremente Y. Implica sólo que los valores grandes de X están *asociados* con valores grandes de Y. Por ejemplo, en la población de niños, el tamaño del vocabulario y el número de caries no están lo bastante correlacionados positivamente, pero con certeza no es cierto que las caries hagan que crezca el vocabulario. En cambio, los valores de estas dos variables tienden a incrementarse conforme el valor de la edad, una tercera variable, se incrementa. Para niños de una edad fija, quizás existe una muy baja correlación entre el número de caries y el tamaño del vocabulario. En suma, asociación (una alta correlación) no es lo mismo que causa.

EJERCICIOS Sección 5.2 (22–36)

22. Un instructor aplicó un corto examen compuesto de dos partes. Para un estudiante seleccionado al azar, sea X = el número de puntos obtenidos en la primera parte y Y = el número de puntos obtenidos en la segunda parte. Suponga que la función de masa de probabilidad conjunta de X y Y se da en la tabla adjunta.

		y			
$p(x, y)$		0	5	10	15
	0	.02	.06	.02	.10
x	5	.04	.15	.20	.10
	10	.01	.15	.14	.01

a. Si la calificación anotada en la libreta de calificaciones es el número total de puntos obtenidos en las dos partes, ¿cuál es la calificación anotada esperada $E(X + Y)$?

b. Si se anota la máxima de las dos calificaciones, ¿cuál es la calificación anotada esperada?

23. La diferencia entre el número de clientes formados en la caja rápida y el número formado en la caja extrarrápida del ejercicio 3 es $X_1 - X_2$. Calcule la diferencia esperada.

24. Seis individuos, incluidos A y B, se sientan alrededor de una mesa circular en una forma completamente al azar. Suponga que los asientos están numerados 1, . . . , 6. Sea X = el número de asiento de A y Y = el número de asiento de B. Si A envía un mensaje escrito alrededor de la mesa a B en la dirección en la cual están más cerca, ¿cuántos individuos (incluidos A y B) esperaría que manipulen el mensaje?

25. Un topógrafo desea delimitar una región cuadrada con longitud de lado L. Sin embargo, debido a un error de medición, delimita en cambio un rectángulo en el cual los lados norte–sur son de longitud X y los lados este–oeste son de longitud Y. Suponga que X y Y son independientes y que cada uno está uniformemente distribuido en el intervalo $[L - A, L + A]$ (donde $0 < A < L$). ¿Cuál es el área esperada del rectángulo resultante?

26. Considere un pequeño transbordador que puede transportar carros y autobuses. La cuota para carros es de \$3 y para autobuses es de \$10. Sean X y Y el número de carros y autobuses, respectivamente, transportados en un solo viaje. Suponga que la distribución conjunta de X y Y aparece en la tabla del ejercicio 7. Calcule el ingreso esperado en un solo viaje.

27. Annie y Alvie quedaron de encontrarse para desayunar entre el mediodía (0:00 p.m.) y 1:00 p.m. Denote la hora de llegada de Annie por X, y la de Alvie por Y y suponga que X y Y son independientes con funciones de densidad de probabilidad

$$f_X(x) = \begin{cases} 3x^2 & 0 \le x \le 1 \\ 0 & \text{de lo contrario} \end{cases}$$

$$f_Y(y) = \begin{cases} 2y & 0 \le y \le 1 \\ 0 & \text{de lo contrario} \end{cases}$$

¿Cuál es la cantidad esperada de tiempo que la que llega primero debe esperar a la otra persona? [*Sugerencia*: $h(X, Y) = |X - Y|$.]

28. Demuestre que si X y Y son variables aleatorias independientes, entonces $E(XY) = E(X) \cdot E(Y)$. Luego aplique esto en el ejercicio 25. [*Sugerencia*: considere el caso continuo con $f(x, y) = f_X(x) \cdot f_Y(y)$.]

29. Calcule el coeficiente de correlación ρ de X y Y del ejemplo 5.16 (ya se calculó la covarianza).

30. **a.** Calcule la covarianza de X y Y en el ejercicio 22.
b. Calcule ρ para X y Y en el mismo ejercicio.

31. **a.** Calcule la covarianza entre X y Y en el ejercicio 9.
b. Calcule el coeficiente de correlación ρ para X y Y.

32. Reconsidere las vidas útiles de los componentes de minicomputadora X y Y como se describe en el ejercicio 12. Determine $E(XY)$. ¿Qué se puede decir sobre $\text{Cov}(X, Y)$ y ρ?

33. Use el resultado del ejercicio 28 para demostrar que cuando X y Y son independientes, $\text{Cov}(X, Y) = \text{Corr}(X, Y) = 0$.

34. **a.** Recordando la definición de σ^2 para una sola variable aleatoria X, escriba una fórmula que sería apropiada para calcular la varianza de una función $h(X, Y)$ de dos variables aleatorias. [*Sugerencia*: recuerde que la varianza es simplemente un valor esperado especial.]
b. Use esta fórmula para calcular la varianza de la calificación anotada $h(X, Y)$ [$= \text{máx}(X, Y)$ en el inciso (b) del ejercicio 22.

35. **a.** Use las reglas de valor esperado para demostrar que $\text{Cov}(aX + b, cY + d) = ac\,\text{Cov}(X, Y)$.
b. Use el inciso (a) junto con las reglas de varianza y desviación estándar para demostrar que $\text{Corr}(aX + b, cY + d) = \text{Corr}(X, Y)$ cuando a y c tienen el mismo signo.
c. ¿Qué sucede si a y c tienen signos opuestos?

36. Demuestre que si $Y = aX + b$ ($a \neq 0$), entonces $\text{Corr}(X, Y) = +1$ o -1. ¿En qué condiciones será $\rho = +1$?

5.3 Estadísticos y sus distribuciones

En el capítulo 1, $x_1, x_2, \ldots, x_n$ denotaron las observaciones en una sola muestra. Considérese seleccionar dos muestras diferentes de tamaño n de la misma distribución de población. Las x_i en la segunda muestra diferirán siempre virtualmente por lo menos un poco de aquellas en la primera muestra. Por ejemplo, una primera muestra de $n = 3$ carros de un tipo particular podría producir eficiencias de combustible $x_1 = 30.7$, $x_2 = 29.4$, $x_3 = 31.1$, mientras que una segunda muestra puede dar $x_1 = 28.8$, $x_2 = 30.0$ y $x_3 = 32.5$. Antes de obtener datos, existe incertidumbre sobre el valor de cada x_i. Debido a esta incertidumbre, *antes* de que los datos estén disponibles, cada observación se considera como una variable aleatoria y la muestra se denota por $X_1, X_2, \ldots, X_n$ (letras mayúsculas para variables aleatorias).

Esta variación en los valores observados implica a su vez que el valor de cualquier función de las observaciones muestrales, tal como la media muestral, la desviación estándar muestral o la dispersión de los cuartos muestrales, también varía de una muestra a otra. Es decir, antes de obtener $x_1, \ldots, x_n$, existe incertidumbre en cuanto al valor de $\bar{x}$, el valor de s, y así sucesivamente.

Ejemplo 5.19 Suponga que la resistencia del material de un espécimen seleccionado al azar de un tipo particular tiene una distribución Weibull con valores de parámetro $\alpha = 2$ (forma) y $\beta = 5$ (escala). La curva de densidad correspondiente se muestra en la figura 5.6. Las fórmulas de la sección 4.5 dan

$$\mu = E(x) = 4.4311 \quad \tilde{\mu} = 4.1628 \quad \sigma^2 = V(X) = 5.365 \quad \sigma = 2.316$$

La media excede a la mediana debido a la asimetría positiva de la distribución.

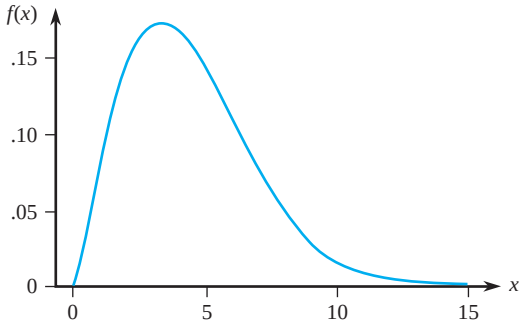


Figura 5.6 Curva de densidad Weibull del ejemplo 5.19

Se utilizó software estadístico para generar seis muestras diferentes, cada una con $n = 10$, de esta distribución (resistencias de material de seis diferentes grupos de diez especímenes cada uno). Los resultados aparecen en la tabla 5.1, seguidos por los valores de la media, la mediana y la desviación estándar de cada muestra. Obsérvese en primer lugar que las diez observaciones en cualquier muestra particular son diferentes de aquellas en cualquier otra muestra. En segundo lugar, los seis valores de la media muestral son diferentes entre sí, como lo son los seis valores de la mediana muestral y los seis valores de la desviación estándar de la muestra. Lo mismo es cierto para las medias recortadas 10%, la dispersiones de los cuartos de las muestras, y así sucesivamente.

Tabla 5.1 Muestras de la distribución Weibull del ejemplo 5.19

Muestra	1	2	3	4	5	6
1	6.1171	5.07611	3.46710	1.55601	3.12372	8.93795
2	4.1600	6.79279	2.71938	4.56941	6.09685	3.92487
3	3.1950	4.43259	5.88129	4.79870	3.41181	8.76202
4	0.6694	8.55752	5.14915	2.49759	1.65409	7.05569
5	1.8552	6.82487	4.99635	2.33267	2.29512	2.30932
6	5.2316	7.39958	5.86887	4.01295	2.12583	5.94195
7	2.7609	2.14755	6.05918	9.08845	3.20938	6.74166
8	10.2185	8.50628	1.80119	3.25728	3.23209	1.75468
9	5.2438	5.49510	4.21994	3.70132	6.84426	4.91827
10	4.5590	4.04525	2.12934	5.50134	4.20694	7.26081
$\bar{x}$	4.401	5.928	4.229	4.132	3.620	5.761
$\tilde{x}$	4.360	6.144	4.608	3.857	3.221	6.342
s	2.642	2.062	1.611	2.124	1.678	2.496

Además, el valor de la media de cualquier muestra puede ser considerado como *estimación puntual* (“puntual” porque es un solo número, correspondiente a un solo punto sobre la línea de numeración) de la media de la población μ, cuyo valor se sabe que es 4.4311. Ninguna de las estimaciones de estas seis muestras es idéntica a la que se está estimando. Las estimaciones de la segunda y sexta muestras son demasiado grandes, en tanto que la quinta da una subestimación sustancial. Asimismo, la desviación estándar muestral da una estimación puntual de la desviación estándar de la población. Las seis estimaciones resultantes están equivocadas por lo menos en una pequeña cantidad.

En resumen, los valores de las observaciones de las muestras individuales varían de una muestra a otra por lo que, en general, el valor de cualquier cantidad calculada a partir de datos de la muestra y el valor de una característica de la muestra utilizada como una estimación de la característica de la población correspondiente casi nunca coinciden con lo que se estimó. ■

DEFINICIÓN

Un **estadístico** es cualquier cantidad cuyo valor puede ser calculado a partir de datos muestrales. Antes de obtener los datos, existe incertidumbre sobre qué valor de cualquier estadístico particular resultará. Por consiguiente, un estadístico es una variable aleatoria y será denotado por una letra mayúscula; para representar el valor calculado u observado del estadístico se utiliza una letra minúscula.

Por lo tanto la media muestral, considerada como estadístico (antes de seleccionar una muestra o realizar un experimento), está denotada por $\overline{X}$; el valor calculado de este estadístico es $\bar{x}$. Del mismo modo, S representa la desviación estándar muestral considerada como estadístico y su valor calculado es s. Si se seleccionan muestras de dos tipos diferentes de ladrillos y las resistencias a la compresión individuales se denotan por $X_1, \ldots, X_m$ y $Y_1, \ldots, Y_n$, respectivamente, entonces el estadístico $\overline{X} - \overline{Y}$, la diferencia entre las dos resistencias a la compresión muestrales medias, a menudo es de gran interés.

Cualquier estadístico, por el hecho de ser una variable aleatoria, tiene una distribución de probabilidad. En particular, la media muestral $\overline{X}$ tiene una distribución de probabilidad. Supóngase, por ejemplo, que $n = 2$ componentes se seleccionan al azar y que el número de descomposturas mientras se encuentran dentro de garantía se determina para cada uno. Los valores posibles del número medio muestral de descomposturas $\overline{X}$ son 0 (si $X_1 = X_2 = 0$), .5 (si $X_1 = 0$ y $X_2 = 1$ o $X_1 = 1$ y $X_2 = 0$), 1, 1.5, . . . La distribución de probabilidad de $\overline{X}$ especifica $P(\overline{X} = 0)$, $P(\overline{X} = .5)$ y así sucesivamente, a partir de las cuales otras probabilidades tales como $P(1 \leq \overline{X} \leq 3)$ y $P(\overline{X} \geq 2.5)$ pueden ser calculadas. Asimismo, si para una muestra de tamaño $n = 2$, los únicos valores posibles de la varianza muestral son 0, 12.5 y 50 (el cual es el caso si X_1 y X_2 pueden tomar sólo los valores 40, 45 o 50), entonces la distribución de probabilidad de S^2 da $P(S^2 = 0)$, $P(S^2 = 12.5)$ y $P(S^2 = 50)$. La distribución de probabilidad de un estadístico en ocasiones se conoce como **distribución de muestreo** para enfatizar que describe cómo varía el valor del estadístico a través de todas las muestras que pudieran ser seleccionadas.

Muestras aleatorias

La distribución de probabilidad de cualquier estadístico particular depende no sólo de la distribución de la población (normal, uniforme, etc.) y el tamaño de muestra n sino también del método de muestreo. Considérese seleccionar una muestra de tamaño $n = 2$ de una población compuesta de sólo los tres valores 1, 5 y 10, y supóngase que el estadístico de interés es la varianza muestral. Si el muestreo se realiza "con reemplazo", entonces $S^2 = 0$ resultará si $X_1 = X_2$. Sin embargo, S^2 no puede ser igual a 0 si el muestreo se realiza "sin reemplazo". Por tanto $P(S^2 = 0) = 0$ con un método de muestreo y esta probabilidad es positiva con el otro método. La siguiente definición describe un método de muestreo encontrado a menudo (por lo menos aproximadamente) en la práctica.

DEFINICIÓN

Se dice que las variables aleatorias $X_1, X_2, \ldots, X_n$ forman una **muestra aleatoria (**simple) de tamaño n si

1. Las X_i son variables aleatorias independientes.

2. Cada X_i tiene la misma distribución de probabilidad.

Las condiciones 1 y 2 pueden ser parafraseadas diciendo que las X_i son *independientes e idénticamente distribuidas* (iid). Si el muestreo se realiza con reemplazo o de una población infinita (conceptual), las condiciones 1 y 2 se satisfacen con exactitud. Estas condiciones serán aproximadamente satisfechas si el muestreo se realiza sin reemplazo, aunque el tamaño n de la muestra es mucho más pequeño que el tamaño de la población N. En la práctica, si $n/N \leq .05$ (cuando mucho 5% de la población se muestrea), se puede proceder

como si las X_i formaran una muestra aleatoria. La virtud de este método de muestreo es que la distribución de probabilidad de cualquier estadístico es más fácil de obtener que con cualquier otro método de muestreo.

Existen dos métodos generales para obtener información sobre una distribución de muestreo de un estadístico. Uno implica cálculos basados en reglas de probabilidad y el otro implica realizar un experimento de simulación.

Deducción de una distribución de muestreo

Se pueden utilizar reglas de probabilidad para obtener la distribución de un estadístico siempre que sea una función "bastante simple" de las X_i y existen relativamente pocos valores X diferentes en la población o bien la distribución de la población tiene una forma "accesible". Los dos ejemplos siguientes ilustran tales situaciones.

Ejemplo 5.20 Una cierta marca de reproductor MP3 viene en tres configuraciones: un modelo con 2 GB de memoria, que cuesta $80, un modelo de 4 GB a un precio de $100 y una versión de 8 GB con un precio de $120. Si el 20% de todos los compradores elige el modelo de 2 GB, el 30% elige el modelo de 4 GB, y el 50% elige el modelo de 8 GB, entonces la distribución de probabilidad del costo X de una sola compra de un reproductor de MP3 seleccionado al azar está dada por

x	80	100	120
$p(x)$	.2	.3	.5

con $\mu = 106$, $\sigma^2 = 244$ (5.2)

Supongamos que en un día particular sólo se venden dos reproductores de MP3. Sean X_1 = los ingresos procedentes de la primera venta y X_2 = los ingresos de la segunda. Supongamos que X_1 y X_2 son independientes, cada uno con la distribución de probabilidad que se muestra en (5.2) [de manera que X_1 y X_2 constituyen una muestra aleatoria de la distribución (5.2)]. En la Tabla 5.2 se enumeran posibles pares (x_1, x_2), la probabilidad de cada uno [calculada utilizando (5.2) y la suposición de independencia] y los valores resultantes de $\bar{x}$ y s^2. [Tenga en cuenta que cuando $n = 2$, $s^2 = (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2$.] Ahora para obtener la distribución de probabilidad $\bar{X}$, la muestra de los ingresos promedio por venta, tenemos que considerar cada posible valor de $\bar{x}$ y calcular su probabilidad. Por ejemplo, $\bar{x} = 100$ aparece tres veces en la tabla con probabilidades de .10, .09 y .10, por lo que

$$p_{\bar{X}}(100) = P(\bar{X} = 100) = .10 + .09 + .10 = .29$$

Asimismo,

$$pS^2(800) = P(S^2 = 800) = P(X_1 = 80, X_2 = 120 \text{ o } X_1 = 120, X_2 = 80)$$
$$= .10 + .10 = .20$$

Tabla 5.2 Resultados, probabilidades y valores de $\bar{x}$ y s^2 en el ejemplo 5.20

x_1	x_2	$p(x_1, x_2)$	$\bar{x}$	s^2
80	80	.04	80	0
80	100	.06	90	200
80	120	.10	100	800
100	80	.06	90	200
100	100	.09	100	0
100	120	.15	110	200
120	80	.10	100	800
120	100	.15	110	200
120	120	.25	120	0

Las distribuciones de muestreo completas de $\bar{X}$ y S^2 aparecen en (5.3) y (5.4).

$\bar{x}$	80	90	100	110	120
$p_{\bar{X}}(\bar{x})$	.04	.12	.29	.30	.25

(5.3)

s^2	0	200	800
$pS^2(s^2)$	.38	.42	.20

(5.4)

La figura 5.7 ilustra un histograma de probabilidad tanto de la distribución original (5.2) como de la distribución $\bar{X}$ (5.3). La figura sugiere primero que la media (valor esperado) de la distribución $\bar{X}$ es igual a la media 106 de la distribución original, puesto que ambos histogramas parecen estar centrados en el mismo lugar.
De acuerdo con (5.3),

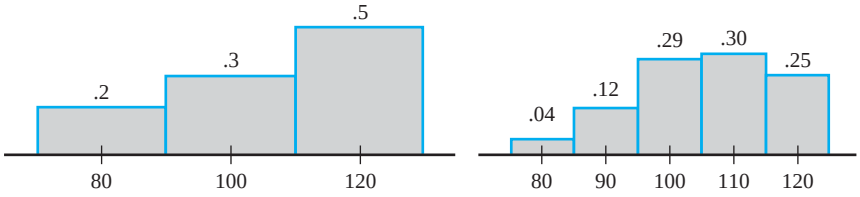


Figura 5.7 Histogramas de probabilidad de la distribución subyacente y distribución $\bar{X}$, en el ejemplo 5.20

$$\mu_{\bar{X}} = E(\bar{X}) = \sum \bar{x} p_{\bar{X}}(\bar{x}) = (80)(.04) + \cdots + (120)(.25) = 106 = \mu$$

En segundo lugar, parece que la distribución $\bar{X}$ tiene una dispersión más pequeña (variabilidad) que la distribución original, puesto que la masa de probabilidad se movió hacia la media. De nuevo de acuerdo con (5.3),

$$\begin{aligned} \sigma^2_{\bar{X}} = V(\bar{X}) &= \sum \bar{x}^2 \cdot p_{\bar{X}}(\bar{x}) - \mu^2_{\bar{X}} \\ &= (80^2)(.04) + \cdots + (120^2)(.25) - (106)^2 \\ &= 122 = \frac{244}{2} = \frac{\sigma^2}{2} \end{aligned}$$

La varianza de $\bar{X}$ es precisamente la mitad de la varianza original (porque $n = 2$).

Utilizando (5.4) el valor medio de S^2 es

$$\begin{aligned} \mu S^2 = E(S^2) &= \sum S^2 \cdot p_{S^2}(s^2) \\ &= (0)(.38) + (200)(.42) + (800)(.20) = 244 = \sigma^2 \end{aligned}$$

Es decir, la distribución de muestreo $\bar{X}$ tiene su centro en la media de la población μ y la distribución de muestreo S^2 está centrada en la varianza de la población σ^2.

Si se hubieran realizado cuatro compras en el día de interés, el ingreso promedio muestral $\bar{X}$ estaría basado en una muestra aleatoria de cuatro X_i, cada una con la distribución (5.2). Más cálculo a la larga da la función de masa de probabilidad de $\bar{X}$ para $n = 4$ como

$\bar{x}$	80	85	90	95	100	105	110	115	120
$p_{\bar{X}}(\bar{x})$	.0016	.0096	.0376	.0936	.1761	.2340	.2350	.1500	.0625

De acuerdo con esto, $\mu_{\bar{X}} = 106 = \mu$ y $\sigma^2_{\bar{X}} = 61 = \sigma^2/4$. La figura 5.8 es un histograma de probabilidad de esta función de masa de probabilidad.

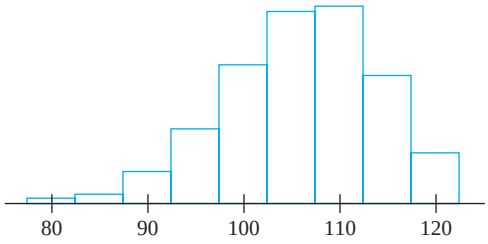


Figura 5.8 Histograma de probabilidad de $\bar{X}$ basado en $n = 4$ en el ejemplo 5.20

El ejemplo 5.20 sugiere primero que todo que los cálculos de $p_{\bar{X}}(\bar{x})$ y $p_{S^2}(s^2)$ pueden ser tediosos. Si la distribución original (5.2) hubiera permitido más de tres valores posibles, entonces incluso con $n = 2$ los cálculos hubieran sido más complicados. El ejemplo también sugiere, sin embargo, que existen algunas relaciones generales entre $E(\bar{X})$, $V(\bar{X})$, $E(S^2)$ y la media μ y la varianza σ^2 de la distribución original. Éstas se formulan en la siguiente sección. Ahora considérese un ejemplo en el cual la muestra aleatoria se saca de una distribución continua.

Ejemplo 5.21 El tiempo de servicio para un tipo de transacción bancaria es una variable aleatoria con distribución exponencial y parámetro λ. Suponga que X_1 y X_2 son tiempos de servicio para dos clientes diferentes, supuestos independientes entre sí. Considere el tiempo de servicio total $T_o = X_1 + X_2$ para los dos clientes, también un estadístico. La función de distribución acumulativa de T_o con $t \geq 0$, es

$$F_{T_o}(t) = P(X_1 + X_2 \leq t) = \iint_{\{(x_1, x_2):\, x_1 + x_2 \leq t\}} f(x_1, x_2)\, dx_1\, dx_2$$

$$= \int_0^t \int_0^{t-x_1} \lambda e^{-\lambda x_1} \cdot \lambda e^{-\lambda x_2}\, dx_2\, dx_1 = \int_0^t [\lambda e^{-\lambda x_1} - \lambda e^{-\lambda t}]\, dx_1$$

$$= 1 - e^{-\lambda t} - \lambda t e^{-\lambda t}$$

La región de integración se ilustra en la figura 5.9.

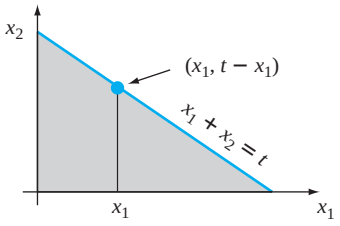


Figura 5.9 Región de integración para obtener la función de distribución acumulativa de T_o en el ejemplo 5.21

La función de densidad de probabilidad de T_o se obtiene diferenciando $F_{T_o}(t)$:

$$f_{T_o}(t) = \begin{cases} \lambda^2 t e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases} \tag{5.5}$$

Ésta es una función de densidad de probabilidad gamma ($\alpha = 2$ y $\beta = 1/\lambda$). La función de densidad de probabilidad de $\bar{X} = T_o/2$ se obtiene a partir de la relación $\{\bar{X} \le \bar{x}\}$ si y sólo si $\{T_o \le 2\bar{x}\}$ como

$$f_{\bar{X}}(\bar{x}) = \begin{cases} 4\lambda^2 \bar{x} e^{-2\lambda \bar{x}} & \bar{x} \ge 0 \\ 0 & \bar{x} < 0 \end{cases} \tag{5.6}$$

La media y la varianza de la distribución exponencial subyacente son $\mu = 1/\lambda$ y $\sigma^2 = 1/\lambda^2$. Con las expresiones (5.5) y (5.6) se puede verificar que $E(\bar{X}) = 1/\lambda$, $V(\bar{X}) = 1/(2\lambda^2)$, $E(T_o) = 2/\lambda$ y $V(T_o) = 2/\lambda^2$. Estos resultados sugieren de nuevo algunas relaciones generales entre medias y varianzas de $\bar{X}$, T_o y la distribución subyacente.

Experimentos de simulación

El segundo método de obtener información sobre distribución de muestreo de un estadístico es realizar un experimento de simulación. Este método casi siempre se utiliza cuando una derivación vía reglas de probabilidad es demasiado difícil o complicada de realizar. Tal experimento virtualmente se realiza siempre con la ayuda de una computadora. Las siguientes características de un experimento deben ser especificadas:

1. El estadístico de interés ($\bar{X}$, S, una media recortada particular, etc.)
2. La distribución de la población (normal con $\mu = 100$ y $\sigma = 15$, uniforme con límite inferior $A = 5$ y superior $B = 10$, etc.)
3. El tamaño de muestra n (p. ej., $n = 10$ o $n = 50$)
4. El número de réplicas k (número de muestras que serán obtenidas)

Luego se utiliza una computadora para obtener k diferentes muestras aleatorias, cada una de tamaño n, de la distribución de población designada. Para cada una de las muestras, calcule el valor del estadístico y construya un histograma de los k valores. Este histograma da la distribución de muestreo *aproximada* del estadístico. Mientras más grande es el valor de k, mejor tenderá a ser la aproximación (la distribución de muestreo real emerge a medida que $k \to \infty$). En la práctica, $k = 500$ o 1000 casi siempre es suficiente si el estadístico es "bastante simple".

Ejemplo 5.22 La distribución de la población del primer estudio de simulación es normal con $\mu = 8.25$ y $\sigma = .75$, como se ilustra en la figura 5.10. [El artículo "Platelet Size in Myocardial Infarction" (*British Med. J.*, 1983: 449–451) sugiere esta distribución de volumen de plaquetas en individuos sin historial clínico de problemas cardiacos serios.]

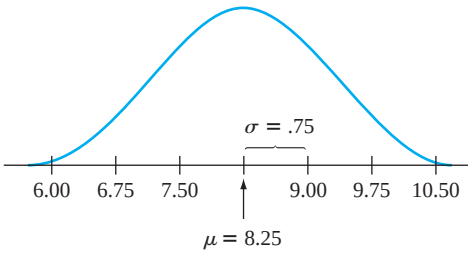


Figura 5.10 Distribución normal, con $\mu = 8.25$ y $\sigma = .75$

En realidad se realizaron cuatro experimentos diferentes, con 500 réplicas por cada uno. En el primero, se generaron 500 muestras de $n = 5$ observaciones cada una con Minitab y los tamaños de las otras tres muestras fueron $n = 10$, $n = 20$ y $n = 30$, respectivamente. La media muestral se calculó para cada muestra, y los histogramas resultantes de valores $\bar{x}$ aparecen en la figura 5.11.

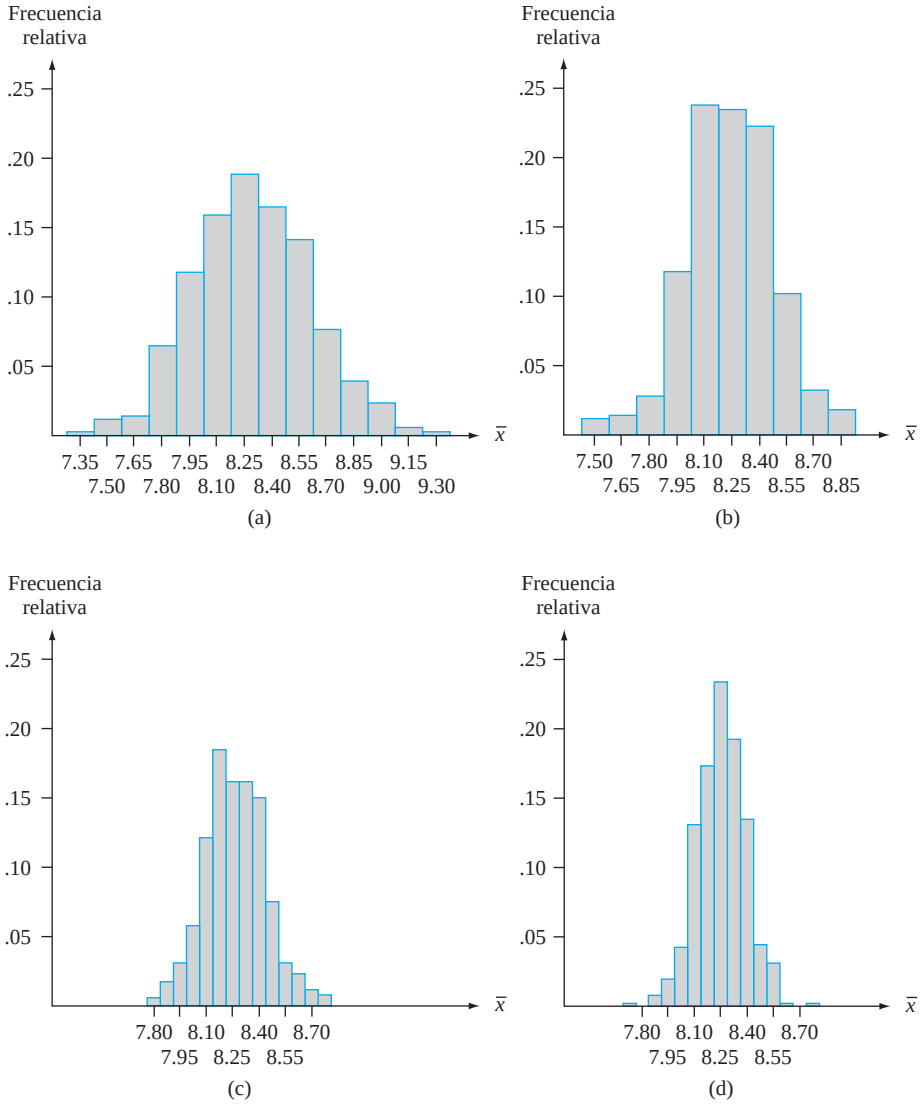


Figura 5.11 Histogramas muestrales de $\bar{x}$ basados en 500 muestras, cada una compuesto de n observaciones: (a) $n = 5$; (b) $n = 10$; (c) $n = 20$; (d) $n = 30$

Lo primero que se nota en relación con los histogramas es su forma. Con una razonable aproximación, cada uno de los cuatro se ve como una curva normal. El parecido sería aún más impactante si cada histograma se hubiera basado en mucho más que 500 valores $\bar{x}$. En segundo lugar, cada histograma está centrado aproximadamente en 8.25, la media de la población muestreada. Si los histogramas se hubieran basado en una secuencia interminable de valores $\bar{x}$, sus centros habrían sido exactamente la media de la población, 8.25.

El aspecto final del histograma es su dispersión uno con respecto al otro. Mientras más grande es el valor de n, más concentrada está la distribución muestral sobre el valor medio. Por eso los histogramas con $n = 20$ y $n = 30$ están basados en intervalos de clase más angostos que aquellos para los dos tamaños de muestra más pequeños. Con los tamaños de muestra más grandes, la mayoría de los valores $\bar{x}$ están bastante cerca de 8.25. Éste es el efecto de promediar. Cuando n es pequeño, un solo valor x inusual puede dar por resultado un valor $\bar{x}$ alejado del centro. Con un tamaño de muestra grande, cualesquiera valores x inusuales, cuando se promedian con los demás valores muestrales, seguirán tendiendo a producir un valor $\bar{x}$ próximo a μ. Si se combinan estas ideas se obtiene un resultado muy apegado a su intuición: **$\bar{X}$ basado en n grande tiende a acercarse más a μ que $\bar{X}$ basado en n pequeño.**

Ejemplo 5.23 Considere un experimento de simulación en el cual la distribución de la población es bastante asimétrica. La figura 5.12 muestra la curva de densidad de las vidas útiles de un tipo de control electrónico [ésta es en realidad una distribución lognormal con $E(\ln(X)) = 3$ y $V(\ln(X)) = .16$]. De nueva cuenta el estadístico de interés es la media muestral $\bar{X}$. El experimento utilizó 500 réplicas y consideró los mismos cuatro tamaños de muestra que en el ejemplo 5.22. Los histogramas resultantes junto con una curva de probabilidad normal generada por Minitab con los 500 valores $\bar{x}$ basados en $n = 30$ se muestran en la figura 5.13.

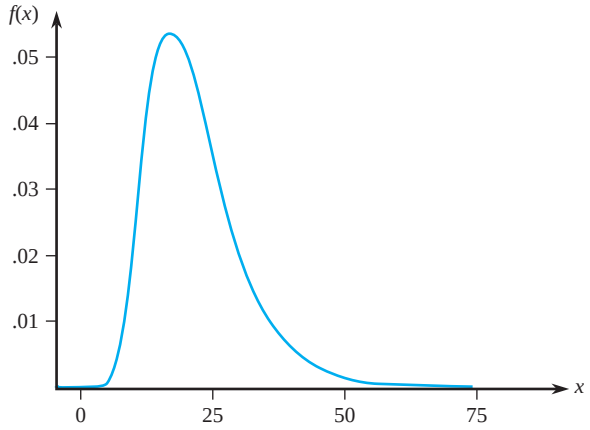
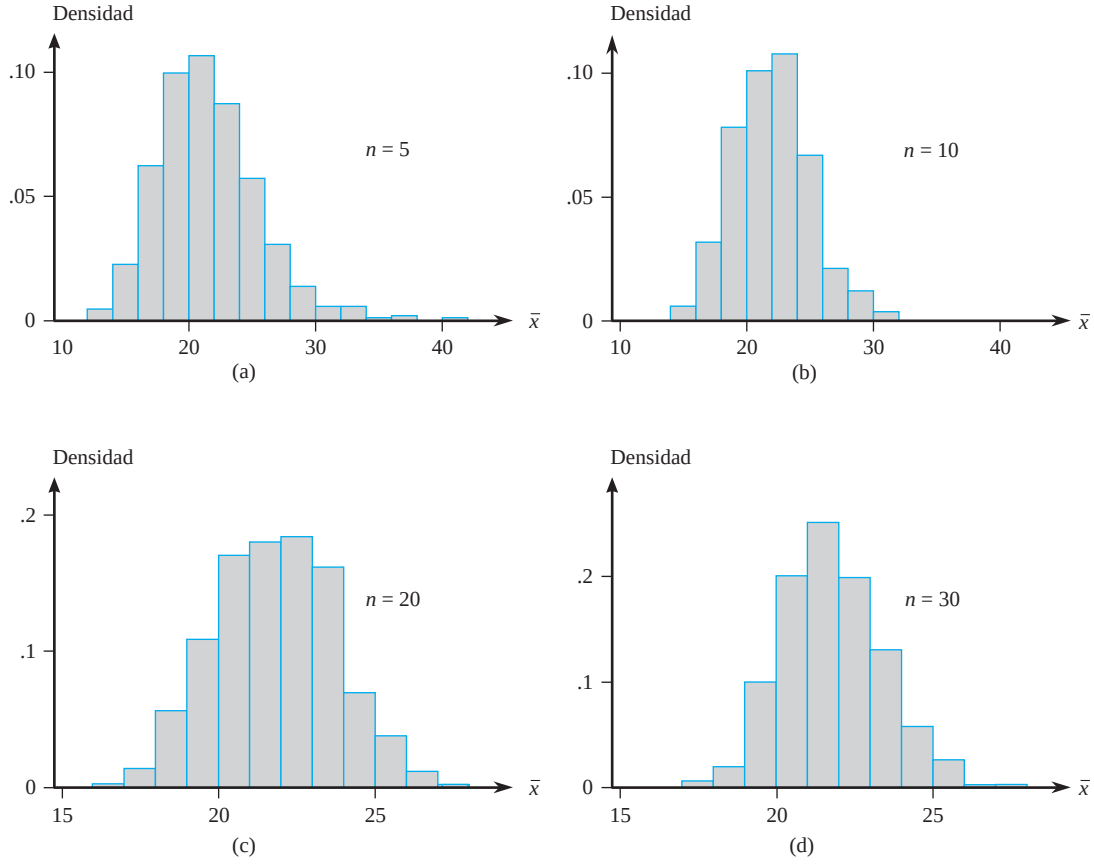
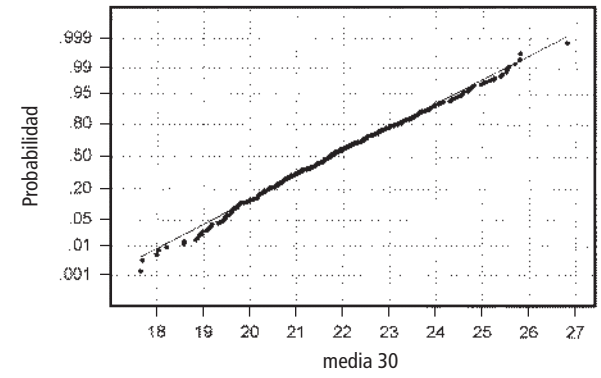


Figura 5.12 Curva de densidad del experimento de simulación del ejemplo 5.23 [$E(X) = 21.7584$, $V(X) = 82.1449$]

A diferencia del caso normal, estos histogramas difieren en cuanto a forma. En particular, se vuelven progresivamente menos asimétricos a medida que el tamaño de muestra n se incrementa. El promedio de los 500 valores $\bar{x}$ con los cuatro tamaños de muestra diferentes se aproximan bastante al valor medio de la distribución de la población. Si cada histograma se hubiera basado en una secuencia interminable de valores $\bar{x}$ en lugar de en sólo 500, los cuatro habrían tenido su centro en exactamente 21.7584. Por tanto, los valores diferentes de n cambian la forma mas no el centro de la distribución de muestreo de $\bar{X}$. La comparación de los cuatro histogramas en la figura 5.13 también muestra que conforme n se incrementa, la dispersión de los histogramas decrece. El incremento de n produce un mayor grado de concentración en torno al valor medio de la población y hace que el histograma se vea más como una curva normal. El histograma de la figura 5.13(d) y la curva de probabilidad normal en la figura 5.13(e) proporcionan una evidencia convincente de que un tamaño de muestra de $n = 30$ es suficiente para superar la asimetría de la distribución de la población y para producir una distribución de muestreo $\bar{X}$ aproximadamente normal.





Promedio: 21.7891
Desv. Est: 1.57396
N: 500

Prueba W de normalidad
R: 0.9975
Valor P (aprox.): > 0.1000

(e)

Figura 5.13 Resultados del experimento de simulación del ejemplo 5.23: (a) histograma de $\bar{x}$ con $n = 5$; (b) histograma de $\bar{x}$ con $n = 10$; (c) histograma de $\bar{x}$ con $n = 20$; (d) histograma de $\bar{x}$ con $n = 30$; (e) curva de probabilidad normal con $n = 30$ (generados por Minitab)

EJERCICIOS Sección 5.3 (37–45)

37. Una marca particular de jabón para lavadora de platos se vende en tres tamaños: 25 oz, 40 oz y 65 oz. Veinte por ciento de todos los compradores seleccionan la caja de 25 oz, 50% seleccionan una caja de 40 oz y el 30% restante seleccionan la caja de 65 oz. Sean X_1 y X_2 los tamaños de paquete seleccionados por dos compradores independientemente seleccionados.

a. Determine la distribución de muestreo de $\overline{X}$, calcule $E(\overline{X})$ y compare con μ.

b. Determine la distribución de muestreo de la varianza muestral S^2, calcule $E(S^2)$ y compare con σ^2.

38. Hay dos semáforos en mi camino de ida y vuelta al trabajo. Sea X_1 el número de semáforos en los cuales me tengo que detener a la ida al trabajo y sea X_2 el número de semáforos en los cuales me tengo que detener de regreso a casa. Suponga que estas dos variables son independientes cada una con una fmp dada en la tabla adjunta (de modo que X_1, X_2 es una muestra aleatoria de tamaño $n = 2$).

x_1	0	1	2
$p(x_1)$	.2	.5	.3

$\mu = 1.1, \sigma^2 = .49$

a. Sea $T_o = X_1 + X_2$ y determine la función de masa de probabilidad.

b. Calcule μ_{T_o}. ¿Cómo se relaciona con μ, la media de la población?

c. Calcule $\sigma^2_{T_o}$. ¿Cómo se relaciona con σ^2, la varianza de la población?

d. Sean X_3 y X_4 el número de luces en las que se requiere una parada durante la conducción hacia el trabajo y de regreso por segundo día consecutivo, suponiendo que es independiente de la primera jornada. Con T_0 = la suma de todos los cuatro X_i, ¿cuáles son ahora los valores de $E(T_0)$ y $V(T_0)$?

e. Refiriéndose de nuevo a (d), ¿cuáles son los valores de $P(T_o = 8)$ y $P(T_o \geq 7)$? [*Sugerencia*: no se te ocurra incluir todos los resultados posibles!]

39. Se sabe que 80% de todos los discos de almacenamiento extraíbles funcionan satisfactoriamente durante el periodo de garantía (son "éxitos"). Suponga que se seleccionan al azar $n = 10$ unidades de disco. Sea X = el número de éxitos en la muestra. El estadístico X/n es la proporción de la muestra (fracción) de éxitos. Obtenga la distribución muestral de este estadístico. [*Sugerencia*: un posible valor de X/n es .3, correspondiente a $X = 3$. ¿Cuál es la probabilidad de este valor (qué clase de variable aleatoria es X)?]

40. Una caja contiene diez sobres sellados numerados $1, \ldots, 10$. Los primeros cinco no contienen dinero, cada uno de los siguientes tres contiene \$5 y hay un billete de \$10 en cada uno de los últimos dos. Se selecciona un tamaño de muestra de 3 *con* reemplazo (así que se tiene una muestra aleatoria) y se obtiene la cantidad más grande en cualquiera de los sobres seleccionados. Si X_1, X_2 y X_3 denotan las cantidades en los sobres seleccionados, el estadístico de interés es M = el máximo de X_1, X_2 y X_3.

a. Obtenga la distribución de probabilidad de este estadístico.

b. Describa cómo realizaría un experimento de simulación para comparar las distribuciones de M con varios tamaños de muestra. ¿Cómo piensa que cambiaría la distribución a medida que n se incrementa?

41. Sea X el número de paquetes enviados por un cliente seleccionado al azar vía una compañía de paquetería y mensajería. Suponga que la distribución de X es como sigue:

x	1	2	3	4
$p(x)$	.4	.3	.2	.1

a. Considere una muestra aleatoria de tamaño $n = 2$ (dos clientes) y sea $\overline{X}$ el número medio muestral de paquetes enviados. Obtenga la distribución de probabilidad de $\overline{X}$.

b. Remítase al inciso (a) y calcule $P(\overline{X} \leq 2.5)$.

c. De nuevo considere una muestra aleatoria de tamaño $n = 2$, pero ahora enfóquese en el estadístico R = el rango muestral (diferencia entre los valores más grande y más pequeño en la muestra). Obtenga la distribución de R. [*Sugerencia*: calcule el valor de R para cada resultado y use las probabilidades del inciso (a).]

d. Si se selecciona una muestra aleatoria de tamaño $n = 4$, ¿cuál es $P(\overline{X} \leq 1.5)$? [*Sugerencia*: no tiene que dar todos los resultados posibles, sólo aquellos para los cuales $\bar{x} \leq 1.5$.]

42. Una compañía mantiene tres oficinas en una región, cada una manejada por dos empleados. Información concerniente a salarios anuales (miles de dólares) es la siguiente:

Oficina	1	1	2	2	3	3
Empleado	1	2	3	4	5	6
Salario	29.7	33.6	30.2	33.6	25.8	29.7

a. Suponga que dos de estos empleados se seleccionan al azar de entre los seis (sin reemplazo). Determine la distribución muestral del salario medio muestral $\overline{X}$.

b. Suponga que se selecciona al azar una de las tres oficinas. Sean X_1 y X_2 los salarios de los dos empleados. Determine la distribución muestral de $\overline{X}$.

c. ¿Cómo se compara $E(\overline{X})$ de los incisos (a) y (b) con el salario medio de la población μ?

43. Suponga que la cantidad de líquido despachada por una máquina está uniformemente distribuida con límite inferior $A = 8$ oz y límite superior $B = 10$ oz. Describa cómo realizaría experimentos de simulación para comparar la distribución muestral de la dispersión de los cuartos (muestral) con tamaños de muestra $n = 5, 10, 20$ y 30.

44. Realice un experimento de simulación con un paquete de computadora estadístico u otro programa para estudiar la distribución muestral de $\overline{X}$ cuando la distribución de la población es de Weibull con $\alpha = 2$ y $\beta = 5$, como en el ejemplo 5.19. Considere los cuatro tamaños de muestra $n = 5, 10, 20$ y 30, y en cada caso utilice 1000 réplicas. ¿Con cuál de estos tamaños de muestra la distribución muestral $\overline{X}$ parece ser aproximadamente normal?

45. Realice un experimento de simulación con un paquete de computadora estadístico u otro programa para estudiar la distribución muestral de $\overline{X}$ cuando la distribución de la población es lognormal con $E(\ln(X)) = 3$ y $V(\ln(X)) = 1$. Considere los cuatro tamaños de muestra $n = 10, 20, 30$ y 50 y en cada caso utilice 1000 réplicas. ¿Con cuál de estos tamaños de muestra la distribución muestral $\overline{X}$ parece ser aproximadamente normal?

5.4 Distribución de la media muestral

La importancia de la media muestral $\overline{X}$ proviene de su uso al sacar conclusiones sobre la media de la población μ. Algunos de los procedimientos inferenciales más frecuentemente utilizados están basados en propiedades de la distribución muestral de $\overline{X}$. Un examen previo de estas propiedades apareció en los cálculos y experimentos de simulación de la sección previa, donde se observaron las relaciones entre $E(\overline{X})$ y μ y también entre $V(\overline{X})$, σ^2 y n.

PROPOSICIÓN

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución con valor medio μ y desviación estándar σ. Entonces

1. $E(\overline{X}) = \mu_{\overline{X}} = \mu$

2. $V(\overline{X}) = \sigma^2_{\overline{X}} = \sigma^2/n$ y $\sigma_{\overline{X}} = \sigma/\sqrt{n}$

Además, con $T_o = X_1 + \cdots + X_n$ (el total de la muestra), $E(T_o) = n\mu$, $V(T_o) = n\sigma^2$ y $\sigma_{T_o} = \sqrt{n}\sigma$.

Las demostraciones de estos resultados se difieren a la siguiente sección. De acuerdo con el resultado 1, la distribución (es decir, probabilidad) muestral de $\overline{X}$ está centrada precisamente en la media de la población de la cual se seleccionó la muestra. El resultado 2 muestra que la distribución $\overline{X}$ se concentra más en torno a μ a medida que se incrementa el tamaño n de la muestra. En un marcado contraste, la distribución de T_o se dispersa más a medida que n se incrementa. Al promediar la probabilidad se mueve hacia la parte media, en tanto que al totalizar la probabilidad se dispersa sobre un rango más y más amplio de valores. La desviación estándar $\sigma_{\overline{X}} = \sigma/\sqrt{n}$ es a menudo llamada *error estándar de la media*; éste describe la magnitud de una desviación típica o representativa de la media muestral respecto de la media poblacional.

Ejemplo 5.24 En una prueba de fatiga por tensión con un espécimen de titanio, el número esperado de ciclos hasta la primera emisión acústica (utilizada para indicar la iniciación del agrietamiento) es $\mu = 28{,}000$, y la desviación estándar del número de ciclos es $\sigma = 5000$. Sea $X_1, X_2, \ldots, X_{25}$ una muestra aleatoria de tamaño 25, donde cada X_i es el número de ciclos en un espécimen diferente seleccionado al azar. Entonces el valor esperado de la media muestral del número de ciclos hasta la primera emisión es $E(\overline{X}) = \mu = 28{,}000$ y el número total esperado de ciclos para los 25 especímenes es $E(T_0) = n\mu = 25(28{,}000) = 700{,}000$. La desviación estándar de $\overline{X}$ (error estándar de la media) y de T_o son

$$\sigma_{\overline{X}} = \sigma/\sqrt{n} = \frac{5000}{\sqrt{25}} = 1000$$

$$\sigma_{T_o} = \sqrt{n}\sigma = \sqrt{25}(5000) = 25{,}000$$

Si el tamaño de la muestra se incrementa a $n = 100$, $E(\overline{X})$ no cambió, pero $\sigma_{\overline{X}} = 500$, la mitad de su valor previo (el tamaño de muestra debe ser cuadruplicado para reducir a la mitad la desviación estándar de $\overline{X}$). ■

El caso de una distribución de población normal

El experimento de simulación del ejemplo 5.22 indicó que cuando la distribución de la población es normal, cada histograma de valores $\bar{x}$ se aproxima muy bien con una curva normal.

PROPOSICIÓN

Sean $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución *normal* con media μ y desviación estándar σ. Entonces con *cualquier* n, $\overline{X}$ está normalmente distribuida (con media μ y desviación estándar $\sigma/\sqrt{n}$), como T_o (con media $n\mu$ y desviación estándar $\sqrt{n}\sigma$).*

Se sabe todo lo que se tiene que saber sobre las distribuciones $\overline{X}$ y T_o cuando la distribución de la población es normal. En particular, probabilidades tales como $P(a \le \overline{X} \le b)$ y $P(c \le T_o \le d)$ se obtienen simplemente estandarizando. La figura 5.14 ilustra la proposición.

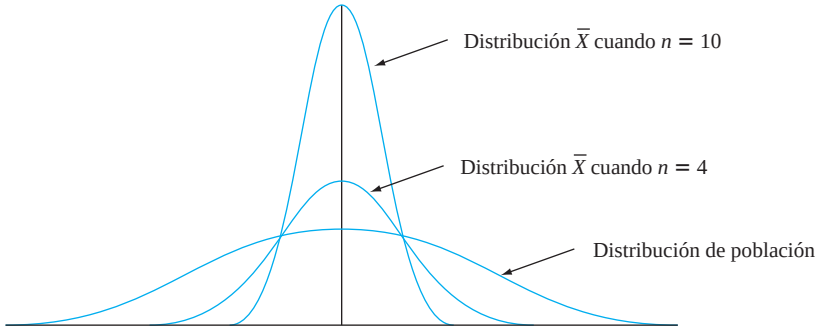


Figura 5.14 Distribución de población normal y distribuciones muestrales $\overline{X}$

Ejemplo 5.25 El tiempo que requiere una rata de cierta subespecie seleccionada al azar para encontrar su camino a través de un laberinto es una variable aleatoria normalmente distribuida con $\mu = 1.5$ min y $\sigma = .35$ min. Suponga que se seleccionan cinco ratas. Sean $X_1, \ldots, X_5$ sus tiempos en el laberinto. Suponiendo que las X_i son una muestra aleatoria de esta distribución normal, ¿cuál es la probabilidad de que el tiempo total $T_o = X_1 + \ldots + X_5$ de las cinco sea de entre 6 y 8 min? De acuerdo con la proposición, T_o tiene una distribución normal con $\mu_{T_o} = n\mu = 5(1.5) = 7.5$ y varianza $\sigma^2_{T_o} = n\sigma^2 = 5(.1225) = .6125$, por tanto $\sigma_{T_o} = .783$. Para estandarizar T_o, reste μ_{T_o} y divida entre σ_{T_o}:

$$P(6 \le T_o \le 8) = P\left(\frac{6 - 7.5}{.783} \le Z \le \frac{8 - 7.5}{.783}\right)$$
$$= P(-1.92 \le Z \le .64) = \Phi(.64) - \Phi(-1.92) = .7115$$

La determinación de la probabilidad de que el tiempo promedio muestral $\overline{X}$ (una variable normalmente distribuida) sea cuando mucho de 2.0 min requiere $\mu_{\overline{X}} = \mu = 1.5$ y $\sigma_{\overline{X}} = \sigma/\sqrt{n} = .35/\sqrt{5} = .1565$. Entonces

$$P(\overline{X} \le 2.0) = P\left(Z \le \frac{2.0 - 1.5}{.1565}\right) = P(Z \le 3.19) = \Phi(3.19) = .9993$$ ■

* Una prueba del resultado para T_o cuando $n = 2$ es posible si se utiliza el método del ejemplo 5.21, pero los detalles son complicados. El resultado general casi siempre se comprueba por medio de una herramienta teórica llamada *función generadora de momentos*. Se puede consultar una de las referencias del capítulo para más información.

Teorema del límite central

Cuando las X_i están normalmente distribuidas, también lo está $\overline{X}$ con cada tamaño de muestra n. Las deducciones del ejemplo 5.20 y el experimento de simulación del ejemplo 5.23, sugieren que incluso cuando la distribución de la población es altamente no normal, el cálculo de promedios produce una distribución más acampanada que la que está siendo muestreada. Una conjetura razonable es que si n es grande, una curva normal apropiada representará de forma más o menos aproximada la distribución real de $\overline{X}$. El planteamiento formal de este resultado es el más importante teorema de probabilidad.

TEOREMA

Teorema del límite central (TLC)

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución con media μ y varianza σ^2. Entonces si n es suficientemente grande, $\overline{X}$ tiene aproximadamente una distribución normal con $\mu_{\overline{X}} = \mu$ y $\sigma^2_{\overline{X}} = \sigma^2/n$, y T_o también tiene aproximadamente una distribución normal con $\mu_{T_o} = n\mu$, $\sigma^2_{T_o} = n\sigma^2$. Mientras más grande es el valor de n, mejor es la aproximación.

La figura 5.15 ilustra el teorema del límite central. De acuerdo con el TLC, cuando n es grande y se desea calcular una probabilidad como $P(a \le \overline{X} \le b)$, lo único que se requiere es "aparentar" que $\overline{X}$ es normal, estandarizarla y utilizar la tabla normal. La respuesta resultante será aproximadamente correcta. Se podría obtener la respuesta correcta determinando primero la distribución de $\overline{X}$, así que el TLC proporciona un atajo verdaderamente impresionante. La comprobación del teorema implica muchas matemáticas avanzadas.

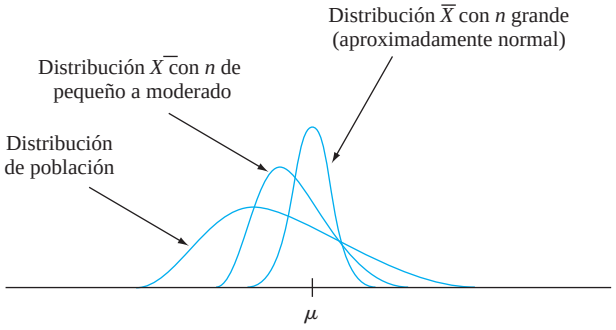


Figura 5.15 Teorema del límite central ilustrado

Ejemplo 5.26 La cantidad de una impureza particular en un lote de cierto producto químico es una variable aleatoria con valor medio de 4.0 g y desviación estándar de 1.5 g. Si se preparan 50 lotes en forma independiente, ¿cuál es la probabilidad (aproximada) de que la cantidad promedio muestral de la impureza $\overline{X}$ esté entre 3.5 a 3.8 g? De acuerdo con la regla empírica que se formulará en breve, $n = 50$ es suficientemente grande como para que el TLC sea aplicable. En ese caso $\overline{X}$ tiene aproximadamente una distribución normal con valor medio $\mu_{\overline{X}} = 4.0$ y $\sigma_{\overline{X}} = 1.5/\sqrt{50} = .2121$, por lo tanto

$$P(3.5 \le \overline{X} \le 3.8) \approx P\left(\frac{3.5 - 4.0}{.2121} \le Z \le \frac{3.8 - 4.0}{.2121}\right)$$
$$= \Phi(-.94) - \Phi(-2.36) = .1645$$

■

Ejemplo 5.27 Una organización de protección al consumidor reporta cotidianamente el número de defectos mayores de cada automóvil nuevo que prueba. Suponga que el número de tales defectos en cierto modelo es una variable aleatoria con valor medio de 3.2 y desviación estándar de 2.4. Entre 100 carros seleccionados al azar de este modelo, ¿qué tan probable es que el número promedio muestral de defectos mayores exceda de 4? Sea X_i el número de defectos mayores del carro *i*ésimo en la muestra aleatoria. Obsérvese que X_i es una variable aleatoria discreta, pero el TLC es aplicable si la variable de interés es discreta o continua. Además, aunque el hecho de que la desviación estándar de esta variable no negativa es bastante grande con respecto al valor medio sugiere que su distribución es positivamente asimétrica, el gran tamaño de muestra implica que $\overline{X}$ sí tiene aproximadamente una distribución normal. Con $\mu_{\overline{X}} = 3.2$ y $\sigma_{\overline{X}} = .24$,

$$P(\overline{X} > 4) \approx P\left(Z > \frac{4 - 3.2}{.24}\right) = 1 - \Phi(3.33) = .0004$$ ■

El TLC da una idea de por qué muchas variables aleatorias tienen distribuciones de probabilidad que son aproximadamente normales. Por ejemplo, el error de medición en un experimento científico puede ser considerado como la suma de varias perturbaciones y errores subyacentes de pequeña magnitud.

Una dificultad práctica al aplicar el teorema del límite central es saber cuándo n es suficientemente grande. El problema es que la precisión de la aproximación con una n particular depende de la forma de la distribución subyacente original que está siendo muestreada. Si la distribución subyacente tiende a una curva de densidad normal, entonces la aproximación será buena incluso con n pequeña, mientras que si está lejos de ser normal, entonces se requerirá una n grande.

> **Regla empírica**
>
> Si $n > 30$, se puede utilizar el teorema del límite central.

Existen distribuciones de población para las cuales incluso una n de 40 o 50 no es suficiente, pero tales distribuciones rara vez se encuentran en la práctica. Por otra parte, la regla empírica a menudo es conservadora; para muchas distribuciones de población, una n mucho menor que 30 sería suficiente. Por ejemplo, en el caso de una distribución de población uniforme, el teorema del límite central da una buena aproximación con $n \geq 12$.

Ejemplo 5.28 Considere la distribución que se muestra en la figura 5.16 para la cantidad comprada (redondeada al dólar más cercano) por un cliente seleccionado al azar en una gasolinera en particular (una distribución similar para las compras en Gran Bretaña (en libras) apareció en el artículo "Data Mining for Fun and Profit", *Statistical Science*, 2000: 111−131; hubo grandes picos en los valores, 10, 15, 20, 25 y 30). La distribución es, obviamente, muy distinta de la normal.

Se le pidió a Minitab seleccionar 1000 muestras diferentes, cada una compuesta de $n = 15$ observaciones y calcular el valor de la media muestral $\overline{X}$ para cada una. La figura 5.17 es un histograma de los 1000 valores resultantes; ésta es la distribución aproximada de x de la toma de muestras en las circunstancias especificadas. Está claro que esta distribución es aproximadamente normal aunque el tamaño de la muestra es en realidad mucho

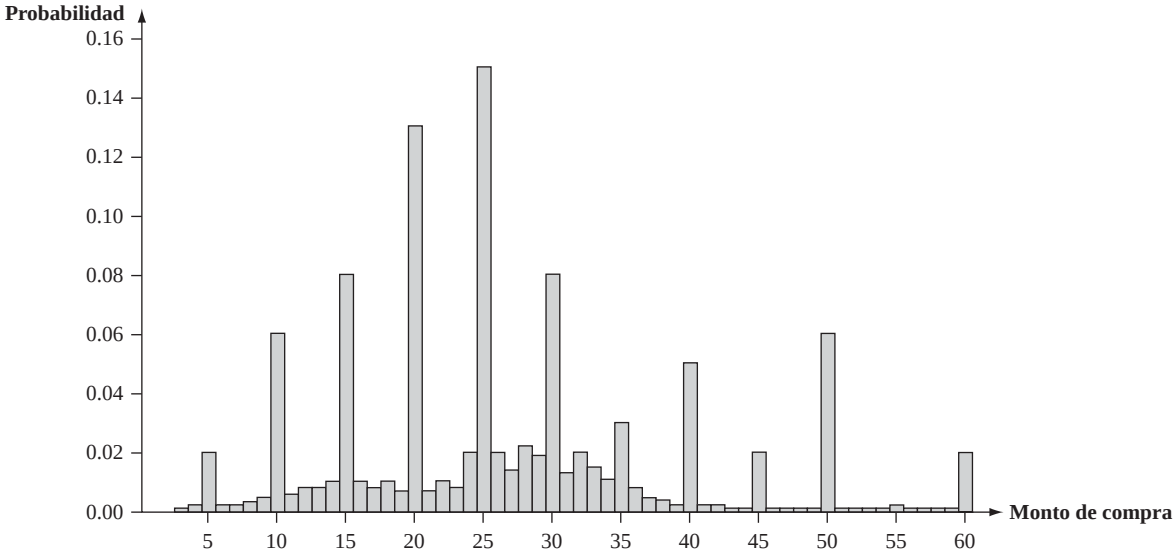


Figura 5.16 Distribución de probabilidad de X = cantidad de gasolina comprada ($)

más pequeña que 30, nuestra regla de oro de corte para invocar el teorema del límite central. Como una prueba más de la normalidad, la figura 5.18 muestra una gráfica de probabilidad normal de los 1000 valores de $\bar{x}$; el patrón lineal es muy prominente. Generalmente no es la no-normalidad en la parte central de la distribución de la población lo que hace que el TLC falle, sino una asimetría muy importante.

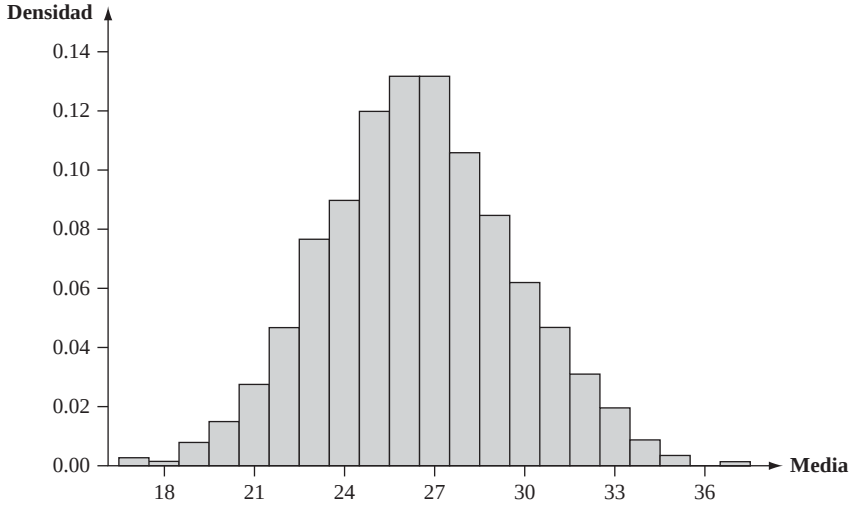


Figura 5.17 Distribución de muestreo aproximada de la media muestral de la cantidad comprada cuando $n = 15$ y la distribución de población es como se ve en la figura 5.16

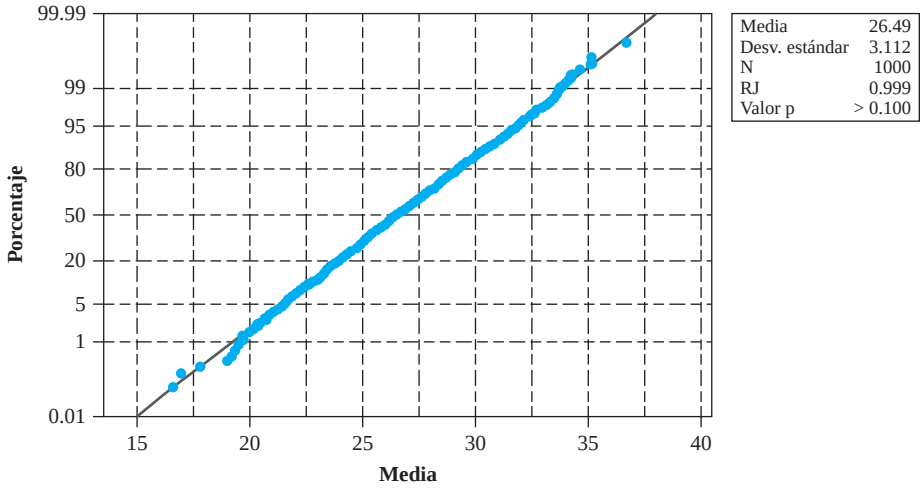


Figura 5.18 Gráfica de la probabilidad normal obtenida con Minitab de los 1000 valores de $\bar{x}$ basada en muestras de tamaño $n = 15$ ■

Otras aplicaciones del teorema del límite central

El teorema del límite central puede ser utilizado para justificar la aproximación normal a la distribución binomial discutida en el capítulo 4. Recuérdese que una variable binomial X es el número de éxitos en una experiencia binomial compuesta de n ensayos independientes con éxitos/fallas y $p = P(S)$ para cualquier ensayo particular. Defina una nueva variable aleatoria X_1 como

$$X_1 = \begin{cases} 1 & \text{si el primer ensayo produce un éxito} \\ 0 & \text{si el primer ensayo produce una falla} \end{cases}$$

y defina $X_2, X_3, \ldots, X_n$ de manera análoga para los otros $n - 1$ ensayos. Cada X_i indica si existe o no un éxito en el ensayo correspondiente.

Como los ensayos son independientes y $P(S)$ es constante de un ensayo a otro, las X_i son independientes e idénticamente distribuidas (una muestra aleatoria de una distribución de Bernoulli). El teorema del límite central implica entonces que si n es suficientemente grande, tanto la suma como el promedio de las X_i tienen distribuciones normales de manera aproximada. Cuando se suman las X_i, se agrega un 1 por cada S (éxito) que ocurra y un 0 por cada F (falla), por tanto $X_1 + \cdots + X_n = X$. La media muestral de las X_i es X/n, la proporción muestral de éxitos. Es decir, tanto X como X/n son aproximadamente normales cuando n es grande. El tamaño de muestra necesario para esta aproximación depende del valor de p: cuando p se acerca a .5, la distribución de cada X_i es razonablemente simétrica (véase la figura 5.19), mientras que la distribución es bastante asimétrica cuando p se acerca a 0 o 1. Utilizando la aproximación sólo si $np \geq 10$ y $n(1 - p) \geq 10$ garantiza que n es suficientemente grande para superar cualquier asimetría en la distribución de Bernoulli subyacente.

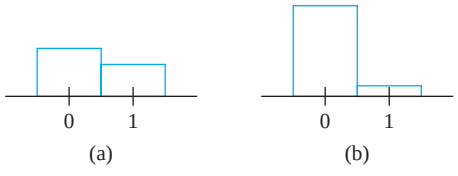


Figura 5.19 Dos distribuciones de Bernoulli: (a) $p = .4$ (razonablemente simétrica); (b) $p = .1$ (muy asimétrica)

Recuérdese de la sección 4.5 que X tiene una distribución lognormal si $\ln(X)$ tiene una distribución normal.

PROPOSICIÓN

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución para la cual sólo son posibles valores positivos $[P(X_i > 0) = 1]$. Entonces si n es suficientemente grande, el producto $Y = X_1X_2 \cdot \cdots \cdot X_n$ tiene aproximadamente una distribución lognormal.

Para verificar esto, obsérvese que

$$\ln(Y) = \ln(X_1) + \ln(X_2) + \cdots + \ln(X_n)$$

Como $\ln(Y)$ es una suma de variables aleatorias independientes y distribuidas de manera idéntica [las $\ln(X_i)$], es aproximadamente normal cuando n es grande, así que Y tiene aproximadamente una distribución lognormal. Como ejemplo de la aplicabilidad de este resultado, Bury (*Statistical Models in Applied Science*, Wiley, pág. 590) argumenta que el proceso de daños en el flujo plástico y en la propagación de grietas es un proceso multiplicativo, de modo que variables tales como el porcentaje de alargamiento y la resistencia a la ruptura tienen aproximadamente distribuciones lognormales.

EJERCICIOS Sección 5.4 (46–57)

46. El diámetro interno de un anillo de pistón seleccionado al azar es una variable aleatoria con valor medio de 12 cm y desviación estándar de .04 cm.

- **a.** Si $\overline{X}$ es el diámetro medio en una muestra aleatoria de $n = 16$ anillos, ¿dónde está centrada la distribución muestral de $\overline{X}$ y cuál es la desviación estándar de la distribución $\overline{X}$?
- **b.** Responda las preguntas planteadas en el inciso (a) con un tamaño de muestra de $n = 64$ anillos.
- **c.** ¿Con cuál de las dos muestras aleatorias, la del inciso (a) o la del inciso (b), es más probable que $\overline{X}$ esté dentro de .01 cm de 12 cm? Explique su razonamiento.

47. Remítase al ejercicio 46. Suponga que la distribución del diámetro es normal.

- **a.** Calcule $P(11.99 \leq \overline{X} \leq 12.01)$, cuando $n = 16$.
- **b.** ¿Qué tan probable es que el diámetro medio muestral exceda de 12.01 cuando $n = 25$?

48. La National Health Statistics Reports en un informe de fecha 22 de octubre de 2008, declaró que para un tamaño de muestra de 277 hombres estadounidenses de 18 años de edad, la media muestral de la circunferencia de la cintura fue de 86.3 cm. Un método algo complicado se utilizó para *estimar* varios percentiles de la población, dando como resultado los siguientes valores:

5º	10º	25º	50º	75º	90º	95º
69.6	70.9	75.2	81.3	95.4	107.1	116.4

- **a.** ¿Es plausible que la distribución de tamaño de la cintura sea por lo menos aproximadamente normal? Explique su razonamiento. Si su respuesta es no, haga una conjetura de la forma de la distribución de la población.
- **b.** Supongamos que la media poblacional del tamaño de la cintura es de 85 cm y que la desviación estándar de la población es de 15 cm. ¿Qué tan probable es que una muestra aleatoria de 277 individuos resulte en una media muestral del tamaño de la cintura de al menos 86.3 cm?
- **c.** Volviendo al inciso (b), supongamos ahora que la media poblacional del tamaño de la cintura es de 82 cm. Ahora, ¿cuál es la probabilidad (aproximada) de que la media de la muestra será de al menos 86.3 cm? A la luz de este cálculo, ¿cree usted que 82 cm es un valor razonable para μ?

49. Hay 40 estudiantes en una clase de estadística elemental. Basado en años de experiencia, el instructor sabe que el tiempo requerido para calificar un primer examen seleccionado al azar es una variable aleatoria con un valor esperado de 6 min y una desviación estándar de 6 min.

- **a.** Si los tiempos de calificación son independientes y el instructor comienza a calificar a las 6:50 p.m., y califica en forma continua, ¿cuál es la probabilidad (aproximada) de que termine de calificar antes de que se inicie el programa de noticias de las 11:00 p.m.?
- **b.** Si el reporte de deportes se inicia a la 11:10, ¿cuál es la probabilidad de que se pierda una parte del reporte si espera hasta que termine de calificar para prender la TV?

50. La resistencia a la ruptura de un remache tiene un valor medio de 10,000 lb/pulg^2 y una desviación estándar de 500 lb/pulg^2.

- **a.** ¿Cuál es la probabilidad de que la resistencia a la ruptura media de una muestra aleatoria de 40 remaches sea de entre 9900 y 10,200?
- **b.** Si el tamaño de muestra hubiera sido de 15 y no de 40, ¿se podría calcular la probabilidad solicitada en el inciso (a) con la información dada?

51. El tiempo requerido por un solicitante de una hipoteca seleccionado al azar para llenar un formulario tiene una distribución normal con valor medio de 10 minutos y desviación estándar de 2 min. Si cinco individuos llenan un formulario en un día y seis en otro, ¿cuál es la probabilidad de que la cantidad de tiempo promedio muestral requerido cada día sea cuando mucho de 11 minutos?

52. La vida útil de un tipo de batería está normalmente distribuida con valor medio de 10 horas y desviación estándar de 1 hora.

Hay cuatro baterías en un paquete. ¿Qué valor de vida útil es tal que la vida útil total de todas las baterías contenidas en un paquete exceda ese valor en sólo 5% de todos los paquetes?

53. Se sabe que la dureza Rockwell de pines de un tipo tiene un valor medio de 50 y una desviación estándar de 1.2.
- **a.** Si la distribución es normal, ¿cuál es la probabilidad de que la dureza media de una muestra aleatoria de 9 pines sea por lo menos de 51?
- **b.** Sin suponer una población normal, ¿cuál es la probabilidad (aproximada) de que la dureza media de una muestra aleatoria de 40 pines sea por lo menos de 51?

54. Suponga que la densidad de un sedimento (g/cm) de un espécimen seleccionado al azar de cierta región está normalmente distribuida con media de 2.65 y desviación estándar de .85 (sugerida en "Modeling Sediment and Water Column Interactions for Hydrophobic Pollutants", *Water Research*, 1984: 1169–1174).
- **a.** Si se selecciona una muestra aleatoria de 25 especímenes, ¿cuál es la probabilidad de que la densidad del sedimento promedio muestral sea cuando mucho de 3.00? ¿De entre 2.65 y 3.00?
- **b.** ¿Qué tan grande debe ser un tamaño de muestra para garantizar que la primera probabilidad en el inciso (a) sea por lo menos de .99?

55. El número de infracciones de estacionamiento aplicadas en una ciudad en cualquier día dado de la semana tiene una distribución de Poisson con parámetro $\mu = 50$. ¿Cuál es la probabilidad aproximada de que
- **a.** entre 35 y 70 infracciones sean aplicadas en un día particular? [*Sugerencia*: cuando μ es grande, una variable aleatoria de Poisson tiene aproximadamente una distribución normal.]
- **b.** el número total de infracciones aplicadas durante una semana de 5 días sea de entre 225 y 275?

56. Un canal de comunicación binaria transmite una secuencia de "bits" (ceros y unos). Suponga que por cualquier bit particular transmitido, existe 10% de probabilidad de que ocurra un error en la transmisión (un 0 se convierte en 1 o un 1 se convierte en 0). Suponga que los errores en los bits ocurren independientemente uno de otro.
- **a.** Considere transmitir 1000 bits. ¿Cuál es la probabilidad aproximada de que cuando mucho ocurran 125 errores de transmisión?
- **b.** Suponga que el mismo mensaje de 1000 bits es enviado en dos momentos diferentes independientemente uno de otro. ¿Cuál es la probabilidad aproximada de que el número de errores en la primera transmisión esté dentro de 50 del número de errores en la segunda?

57. Suponga que la distribución del tiempo X (en horas) utilizado por estudiantes en cierta universidad en un proyecto particular es gamma con parámetros $\alpha = 50$ y $\beta = 2$. Como α es grande, se puede demostrar que X tiene aproximadamente una distribución normal. Use este hecho para calcular la probabilidad aproximada de que un estudiante seleccionado al azar utilice cuando mucho 125 horas en el proyecto.

5.5 Distribución de una combinación lineal

La media muestral $\overline{X}$ y el total muestral T_o son casos especiales de un tipo de variable aleatoria que surgen con frecuencia en aplicaciones estadísticas.

DEFINICIÓN

Dado un conjunto de n variables aleatorias $X_1, \ldots, X_n$ y n constantes numéricas $a_1, \ldots, a_n$, la variable aleatoria

$$Y = a_1X_1 + \cdots + a_nX_n = \sum_{i=1}^{n} a_iX_i \qquad (5.7)$$

se llama **combinación lineal** de las X_i.

Por ejemplo, $4X_1 - 5X_2 + 8X_3$ es una combinación lineal de X_1, X_2 y X_3 con $a_1 = 4$; $a_2 = -5$ y $a_3 = 8$.

Tomando $a_1 = a_2 = \cdots = a_n = 1$ da como resultado $Y = X_1 + \cdots + X_n = T_o$ y $a_1 = a_2 = \cdots = a_n = \frac{1}{n}$ resulta en

$$Y = \frac{1}{n}X_1 + \cdots + \frac{1}{n}X_n = \frac{1}{n}(X_1 + \cdots + X_n) = \frac{1}{n}T_o = \overline{X}$$

Obsérvese que no se requiere que las X_i sean independientes o estén idénticamente distribuidas. Todas las X_i podrían tener distribuciones diferentes y por consiguiente valores medios y varianzas diferentes. Primero se considera el valor esperado y la varianza de una combinación lineal.

PROPOSICIÓN

Sean $X_1, X_2, \ldots, X_n$ con valores medios $\mu_1, \ldots, \mu_n$, respectivamente, y varianzas $\sigma_1^2, \ldots, \sigma_n^2$, respectivamente.

1. Si las X_i son independientes o no

$$\begin{aligned} E(a_1X_1 + a_2X_2 + \cdots + a_nX_n) &= a_1E(X_1) + a_2E(X_2) + \cdots + a_nE(X_n) \\ &= a_1\mu_1 + \cdots + a_n\mu_n \end{aligned} \tag{5.8}$$

2. Si $X_1, \ldots, X_n$ son independientes,

$$\begin{aligned} V(a_1X_1 + a_2X_2 + \cdots + a_nX_n) &= a_1^2V(X_1) + a_2^2V(X_2) + \cdots + a_n^2V(X_n) \\ &= a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2 \end{aligned} \tag{5.9}$$

y

$$\sigma_{a_1X_1+\cdots+a_nX_n} = \sqrt{a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2} \tag{5.10}$$

3. Con cualquier $X_1, \ldots, X_n$,

$$V(a_1X_1 + \cdots + a_nX_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_ia_j\mathrm{Cov}(X_i, X_j) \tag{5.11}$$

Las comprobaciones se dan al final de la sección. Un parafraseo de (5.8) es que el valor esperado de una combinación lineal es la misma combinación lineal de los valores esperados; por ejemplo, $E(2X_1 + 5X_2) = 2\mu_1 + 5\mu_2$. El resultado (5.9) en la proposición 2 es un caso especial de (5.11) en la proposición 3; cuando las X_i son independientes, $\mathrm{Cov}(X_i, X_j) = 0$ para $i \neq j$ y $= V(X_i)$ para $i = j$ (esta simplificación en realidad ocurre cuando las X_i no están correlacionadas, una condición más débil que la de independencia). Especializando al caso de una muestra aleatoria (X_i independientes e idénticamente distribuidas) con $a_i = 1/n$ para cada i da $E(\bar{X}) = \mu$ y $V(\bar{X}) = \sigma^2/n$ como se discutió en la sección 5.4. Un comentario similar se aplica a las reglas para T_o.

Ejemplo 5.29

Una gasolinera vende tres grados de gasolina; regular, extra y súper. Éstas se venden a \$3.00, \$3.20 y \$3.40 por galón, respectivamente. Sean X_1, X_2 y X_3 las cantidades (galones) de estas gasolinas compradas en un día particular. Suponga que las X_i son independientes con $\mu_1 = 1000$, $\mu_2 = 500$, $\mu_3 = 300$, $\sigma_1 = 100$, $\sigma_2 = 80$ y $\sigma_3 = 50$. El ingreso por las ventas es $Y = 3.0X_1 + 3.2X_2 + 3.4X_3$ y

$$\begin{aligned} E(Y) &= 3.0\mu_1 + 3.2\mu_2 + 3.4\mu_3 = \$5620 \\ V(Y) &= (3.0)^2\sigma_1^2 + (3.2)^2\sigma_2^2 + (3.4)^2\sigma_3^2 = 184{,}436 \\ \sigma_Y &= \sqrt{184{,}436} = \$429.46 \end{aligned}$$

■

Diferencia entre dos variables aleatorias

Un importante caso especial de una combinación lineal se presenta con $n = 2$, $a_1 = 1$ y $a_2 = -1$:

$$Y = a_1X_1 + a_2X_2 = X_1 - X_2$$

Entonces se tiene el siguiente corolario de la proposición.

COROLARIO

$E(X_1 - X_2) = E(X_1) - E(X_2)$ para dos variables aleatorias cualesquiera X_1 y X_2.
$V(X_1 - X_2) = V(X_1) + V(X_2)$ si X_1 y X_2 son variables aleatorias independientes.

El valor esperado de una diferencia es la diferencia de los dos valores esperados, pero la varianza de una diferencia entre dos variables independientes es la *suma*, *no* la diferencia, de las dos varianzas. Existe tanta variabilidad en $X_1 - X_2$ como en $X_1 + X_2$ [escribiendo $X_1 - X_2 = X_1 + (-1)X_2$; $(-1)X_2$ tiene la misma cantidad de variabilidad que X_2].

Ejemplo 5.30 Una compañía automotriz equipa un modelo particular con un motor de seis cilindros o un motor de cuatro cilindros. Sean X_1 y X_2 eficiencias de combustible de automóviles de seis y cuatro cilindros seleccionados en forma independiente al azar, respectivamente. Con $\mu_1 = 22$, $\mu_2 = 26$, $\sigma_1 = 1.2$ y $\sigma_2 = 1.5$,

$$E(X_1 - X_2) = \mu_1 - \mu_2 = 22 - 26 = -4$$
$$V(X_1 - X_2) = \sigma_1^2 + \sigma_2^2 = (1.2)^2 + (1.5)^2 = 3.69$$
$$\sigma_{X_1 - X_2} = \sqrt{3.69} = 1.92$$

Si se cambia la notación de modo que X_1 se refiera al automóvil de cuatro cilindros, entonces $E(X_1 - X_2) = 4$, pero la varianza de la diferencia sigue siendo de 3.69. ■

El caso de variables aleatorias normales

Cuando las X_i forman una muestra aleatoria de una distribución normal, $\overline{X}$ y T_o están normalmente distribuidas. He aquí un resultado más general con respecto a combinaciones lineales.

PROPOSICIÓN

Si $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes normalmente distribuidas (con quizá diferentes medias y/o varianzas), entonces cualquier combinación lineal de las X_i también tiene una distribución normal. En particular, la diferencia $X_1 - X_2$ entre dos variables independientes normalmente distribuidas también está distribuida en forma normal.

Ejemplo 5.31 (Continuación del ejemplo 5.29) El ingreso total por la venta de los tres grados de gasolina en un día particular fue $Y = 3.0X_1 + 3.2X_2 + 3.4X_3$ y se calculó $\mu_Y = 5620$ y (suponiendo independencia) $\sigma_Y = 429.46$. Si las X_i están normalmente distribuidas, la probabilidad de que el ingreso sea de más de 4500 es

$$P(Y > 4500) = P\left(Z > \frac{4500 - 5620}{429.46}\right)$$
$$= P(Z > -2.61) = 1 - \Phi(-2.61) = .9955$$ ■

El teorema del límite central también puede ser generalizado para aplicarlo a ciertas combinaciones lineales. En general, si n es grande y no es probable que algún término individual contribuya demasiado al valor total, entonces Y tiene aproximadamente una distribución normal.

Comprobaciones en el caso $n = 2$

En cuanto al resultado por lo concerniente a los valores esperados, suponga que X_1 y X_2 son continuas con función de densidad de probabilidad conjunta $f(x_1, x_2)$. Entonces

$$\begin{aligned} E(a_1X_1 + a_2X_2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(a_1x_1 + a_2x_2)f(x_1, x_2)\,dx_1\,dx_2 \\ &= a_1\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_1f(x_1, x_2)\,dx_2\,dx_1 \\ &\quad + a_2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_2f(x_1, x_2)\,dx_1\,dx_2 \\ &= a_1\int_{-\infty}^{\infty}x_1 f_{X_1}(x_1)\,dx_1 + a_2\int_{-\infty}^{\infty}x_2 f_{X_2}(x_2)\,dx_2 \\ &= a_1E(X_1) + a_2E(X_2) \end{aligned}$$

La suma reemplaza a la integración en el caso discreto. El argumento en cuanto a la varianza resultante no requiere especificar si la variable es discreta o continua. Recordando que $V(Y) = E[(Y - \mu_Y)^2]$,

$$\begin{aligned} V(a_1X_1 + a_2X_2) &= E\{[a_1X_1 + a_2X_2 - (a_1\mu_1 + a_2\mu_2)]^2\} \\ &= E\{a_1^2(X_1 - \mu_1)^2 + a_2^2(X_2 - \mu_2)^2 + 2a_1a_2(X_1 - \mu_1)(X_2 - \mu_2)\} \end{aligned}$$

La expresión dentro de los paréntesis rectangulares es una combinación lineal de las variables $Y_1 = (X_1 - \mu_1)^2$, $Y_2 = (X_2 - \mu_2)^2$ y $Y_3 = (X_1 - \mu_1)(X_2 - \mu_2)$, así que si se acarrea la operación E a través de los tres términos se obtiene $a_1^2V(X_1) + a_2^2V(X_2) + 2a_1a_2\,\mathrm{Cov}(X_1, X_2)$ como se requiere. ■

EJERCICIOS Sección 5.5 (58–74)

58. Una compañía naviera maneja contenedores en tres diferentes tamaños: (1) 27 pies^3 ($3 \times 3 \times 3$), (2) 125 pies^3 y (3) 512 pies^3. Sea X_i ($i = 1, 2, 3$) el número de contenedores de tipo i embarcados durante una semana dada. Con $\mu_i = E(X_i)$ y $\sigma_i^2 = V(X_i)$, suponga que los valores medios y las desviaciones estándar son como sigue.

$\mu_1 = 200$ $\quad \mu_2 = 250$ $\quad \mu_3 = 100$
$\sigma_1 = 10$ $\quad \sigma_2 = 12$ $\quad \sigma_3 = 8$

a. Suponiendo que X_1, X_2, X_3 son independientes, calcule el valor esperado y la varianza del volumen total embarcado. [*Sugerencia*: volumen $= 27X_1 + 125X_2 + 512X_3$.]

b. ¿Serían sus cálculos necesariamente correctos si las X_i no fueran independientes? Explique.

59. Sean X_1, X_2 y X_3 que representan los tiempos necesarios para realizar tres tareas de reparación sucesivas en cierto taller de servicio. Suponga que son variables aleatorias normales independientes con valores esperados μ_1, μ_2 y μ_3 y varianzas σ_1^2, σ_2^2 y σ_3^2, respectivamente.

a. Si $\mu_1 = \mu_2 = \mu_3 = 60$ y $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 15$ calcule $P(T_o \le 200)$. ¿Cuál es $P(150 \le T_o \le 200)$?

b. Con las μ_i y σ_i dadas en el inciso (a), calcule $P(55 \le \overline{X})$ y $P(58 \le \overline{X} \le 62)$.

c. Con las μ_i y σ_i dadas en el inciso (a), calcule e interprete $P(-10 \le X_1 - .5X_2 - .5X_3 \le 5)$.

d. Si $\mu_1 = 40$, $\mu_2 = 50$, $\mu_3 = 60$, $\sigma_1^2 = 10$, $\sigma_2^2 = 12$ y $\sigma_3^2 = 14$, calcule $P(X_1 + X_2 + X_3 \le 160)$ y $P(X_1 + X_2 \ge 2X_3)$.

60. Cinco automóviles del mismo tipo tienen que realizar un viaje de 300 millas. Los primeros dos utilizarán una marca económica de gasolina y los otros tres una marca de renombre. Sean X_1, X_2, X_3, X_4 y X_5 las eficiencias de combustible observadas (mpg) de los cinco carros. Suponga que estas variables son independientes y normalmente distribuidas con $\mu_1 = \mu_2 = 20$, $\mu_3 = \mu_4 = \mu_5 = 21$ y $\sigma^2 = 4$ con la marca económica y 3.5 con la marca de renombre. Defina una variable aleatoria Y como

$$Y = \frac{X_1 + X_2}{2} - \frac{X_3 + X_4 + X_5}{3}$$

de modo que Y mide la diferencia de eficiencia entre la gasolina económica y la de renombre. Calcule $P(0 \le Y)$ y $P(-1 \le Y \le 1)$. [*Sugerencia*: $Y = a_1X_1 + \cdots + a_5X_5$, con $a_1 = \frac{1}{2}, \ldots, a_5 = -\frac{1}{3}$.]

61. El ejercicio 26 introdujo variables aleatorias X y Y, el número de carros y autobuses, respectivamente, transportados por un transbordador en un solo viaje. La función de masa de probabilidad conjunta de X y Y se da en la tabla del ejercicio 7. Es fácil verificar que X y Y son independientes.

a. Calcule el valor esperado, la varianza y la desviación estándar del número total de vehículos en un solo viaje.

b. Si a cada carro se le cobran \$3 y a cada autobús \$10, calcule el valor esperado, la varianza y la desviación estándar del ingreso resultante de un solo viaje.

62. Un fabricante de cierto componente requiere tres operaciones de maquinado diferentes. El tiempo de maquinado de cada operación tiene una distribución normal y los tres tiempos son

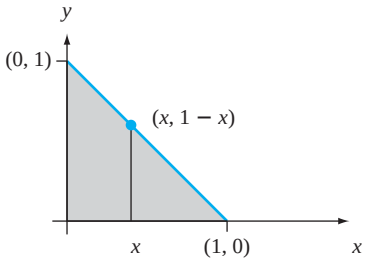
independientes entre sí. Los valores medios son 15, 30 y 20 min, respectivamente, y las desviaciones estándar son 1, 2 y 1.5 min, respectivamente. ¿Cuál es la probabilidad de que se requiera cuando mucho 1 hora de tiempo de maquinado para producir un componente seleccionado al azar?

63. Remítase al ejercicio 3.
- **a.** Calcule la covarianza entre X_1 = el número de clientes en la caja rápida y X_2 = el número de clientes en la caja extrarrápida.
- **b.** Calcule $V(X_1 + X_2)$. ¿Cómo se compara con $V(X_1) + V(X_2)$?

64. Suponga que el tiempo de espera para un autobús en la mañana está uniformemente distribuido en [0, 8], mientras que el tiempo de espera en la noche está uniformemente distribuido en [0, 10] independiente del tiempo de espera en la mañana.
- **a.** Si toma el autobús en la mañana y en la noche durante una semana, ¿cuál es su tiempo de espera total esperado? [*Sugerencia*: defina las variables aleatorias $X_1, \ldots, X_{10}$ y use una regla de valor esperado.]
- **b.** ¿Cuál es la varianza de su tiempo de espera total?
- **c.** ¿Cuáles son el valor esperado y la varianza de la diferencia entre los tiempos de espera en la mañana y en la noche en un día dado?
- **d.** ¿Cuáles son el valor esperado y la varianza de la diferencia entre el tiempo de espera total en la mañana y el tiempo de espera total en la noche durante una semana particular?

65. Suponga que cuando el pH de cierto compuesto químico es 5.00, el pH medido por un estudiante de química de primer año seleccionado al azar es una variable aleatoria con media de 5.00 y desviación estándar de .2. Un gran lote del compuesto se subdivide y a cada estudiante se le da una muestra en un laboratorio matutino y a cada estudiante en un laboratorio vespertino. Sea $\overline{X}$ = el pH promedio determinado por los estudiantes matutinos y $\overline{Y}$ = el pH promedio determinado por los estudiantes vespertinos.
- **a.** Si el pH es una variable normal y hay 25 estudiantes en cada laboratorio, calcule $P(-.1 \le \overline{X} - \overline{Y} \le .1)$. [*Sugerencia*: $\overline{X} - \overline{Y}$ es una combinación lineal de variables normales, así que está normalmente distribuida. Calcule $\mu_{\overline{X}-\overline{Y}}$ y $\sigma_{\overline{X}-\overline{Y}}$.]
- **b.** Si hay 36 estudiantes en cada laboratorio, pero las determinaciones del pH no se suponen normales, calcule (aproximadamente) $P(-.1 \le \overline{X} - \overline{Y} \le .1)$.

66. Si se aplican dos cargas a una viga en voladizo como se muestra en la figura adjunta, el momento de flexión en 0 debido a las cargas es $a_1X_1 + a_2X_2$.



- **a.** Suponga que X_1 y X_2 son variables aleatorias independientes con medias de 2 y 4 kips, respectivamente, y desviaciones estándar de .5 y 1.0 kip, respectivamente. Si $a_1 = 5$ pies y $a_2 = 10$ pies, ¿cuál es el momento de flexión esperado y cuál es la desviación estándar del momento de flexión?
- **b.** Si X_1 y X_2 están normalmente distribuidas, ¿cuál es la probabilidad de que el momento de flexión sea de más de 75 kips-pie?
- **c.** Suponga que las posiciones de las dos cargas son variables aleatorias. Denotándolas por A_1 y A_2, suponga que estas variables tienen medias de 5 y 10 pies, respectivamente, que cada una tiene una desviación estándar de .5, y que todas las A_i y X_i son independientes entre sí. ¿Cuál es el momento esperado ahora?
- **d.** En la situación del inciso (c), ¿cuál es la varianza del momento de flexión?
- **e.** Si la situación es como se describe en el inciso (a) excepto que $\text{Corr}(X_1, X_2) = .5$ (de modo que las dos cargas no sean independientes), ¿cuál es la varianza del momento de flexión?

67. Un tramo de tubería de PVC tiene que ser insertado en otro tramo. La longitud del primer tramo está normalmente distribuida con valor medio de 20 pulg y desviación estándar de .5 pulg. La longitud del segundo tramo es una variable aleatoria normal con media y desviación estándar de 15 pulg y .4 pulg, respectivamente. La cantidad de traslape está normalmente distribuida con valor medio de 1 pulg y desviación estándar de .1 pulg. Suponiendo que los tramos y cantidad de traslape son independientes entre sí, ¿cuál es la probabilidad de que la longitud total después de la inserción sea de entre 34.5 pulg y 35 pulg?

68. Dos aviones vuelan en la misma dirección en dos corredores paralelos adyacentes. En el instante $t = 0$, el primer avión está a 10 km adelante del segundo. Suponga que la velocidad del primer avión (km/h) está normalmente distribuida con media de 520 y desviación estándar de 10 y que la velocidad del segundo también está normalmente distribuida con media y desviación estándar de 500 y 10, respectivamente.
- **a.** ¿Cuál es la probabilidad de que después de 2 horas de vuelo el segundo avión no haya alcanzado al primer avión?
- **b.** Determine la probabilidad de que los aviones estén separados cuando mucho 10 km después de 2 horas.

69. Tres carreteras diferentes entroncan en la entrada de una autopista particular. Suponga que durante un tiempo fijo, el número de carros que llegan por cada carretera a la autopista es una variable aleatoria con valor esperado y desviación estándar como se dan en la tabla

	Carretera 1	Carretera 2	Carretera 3
Valor esperado	800	1000	600
Desviación estándar	16	25	18

- **a.** ¿Cuál es el número de carros total esperado que entran a la autopista en este punto durante el periodo? [*Sugerencia*: sea X_i = el número de la carretera i.]
- **b.** ¿Cuál es la varianza del número total de carros que entran? ¿Ha hecho suposiciones sobre la relación entre los números de carros en las diferentes carreteras?
- **c.** Con X_i denotando el número de carros que entran de la carretera i durante el periodo, suponga que $\text{Cov}(X_1, X_2) = 80$, $\text{Cov}(X_1, X_3) = 90$, y $\text{Cov}(X_2, X_3) = 100$ (de modo que

las tres corrientes de tráfico no son independientes). Calcule el número total esperado de los carros que entran y la desviación estándar del total.

70. Considere una muestra aleatoria de tamaño n tomada de una distribución continua con mediana 0 de modo que la probabilidad de que cualquier observación sea positiva es de .5. Haciendo caso omiso de los signos de las observaciones, clasifíquelas desde las más pequeña a la más grande en valor absoluto y sea W = la suma de los renglones de las observaciones con signos positivos. Por ejemplo, si las observaciones son $-.3$, $+.7$, $+2.1$ y -2.5, entonces los renglones de observaciones positivas son 2 y 3, de modo que $W = 5$. En el capítulo 15, W se llamará *estadístico de filas con signo de Wilcoxon*. W puede representarse como sigue:

$$W = 1 \cdot Y_1 + 2 \cdot Y_2 + 3 \cdot Y_3 + \cdots + n \cdot Y_n = \sum_{i=1}^{n} i \cdot Y_i$$

donde las Y_i son variables aleatorias independientes de Bernoulli, cada una con $p = .5$ ($Y_i = 1$ corresponde a la observación con fila i positiva).

a. Determine $E(Y_i)$ y luego $E(W)$ utilizando la ecuación para W. [*Sugerencia*: los primeros n enteros positivos se suman a $n(n + 1)/2$.]

b. Determine $V(Y_i)$ y luego $V(W)$. [*Sugerencia*: la suma de los cuadrados de los primeros n enteros positivos puede expresarse como $n(n + 1)(2n + 1)/6$.]

71. En el ejercicio 66, el peso de la viga contribuye al momento de flexión. Suponga que la viga es de espesor y densidad uniformes, de modo que la carga resultante esté uniformemente distribuida en la viga. Si el peso de ésta es aleatorio, la carga resultante a consecuencia del peso también es aleatoria; denote esta carga por W (kip-pies).

a. Si la viga es de 12 pies de largo, W tiene una media de 1.5 y una desviación estándar de .25 y las cargas fijas son como se describen en el inciso (a) del ejercicio 66, ¿cuáles son el valor esperado y la varianza del momento de flexión? [*Sugerencia*: si la carga originada por la viga fuera w kip-pies, la contribución al momento de flexión sería $w\int_0^{12} x\,dx$.]

b. Si las tres variables (X_1, X_2 y W) están normalmente distribuidas, ¿cuál es la probabilidad de que el momento de flexión será cuando mucho de 200 kip-pies?

72. Tengo tres encargos que atender en el Edificio de Administración. Sea X_i = el tiempo que requiere el encargo i-ésimo ($i = 1, 2, 3$) y sea X_4 = el tiempo total en minutos que me paso caminando hasta el edificio y de regreso, y entre cada encargo. Suponga que las X_i son independientes y normalmente distribuidas con las siguientes medias y desviaciones estándar: $\mu_1 = 15$, $\sigma_1 = 4$, $\mu_2 = 5$, $\sigma_2 = 1$, $\mu_3 = 8$, $\sigma_3 = 2$, $\mu_4 = 12$, $\sigma_4 = 3$. Pienso salir de mi oficina precisamente a las 10:00 a.m. y deseo pegar una nota en la puerta que diga "Regreso alrededor de las t a.m.". ¿Qué hora debo escribir si deseo que la probabilidad de mi llegada después de t sea de .01?

73. Suponga que la resistencia a la tensión esperada de acero tipo A es de 105 kg/pulg^2 y que la desviación estándar de la resistencia a la tensión es de 8 kg/pulg^2. Para acero tipo B, suponga que la resistencia a la tensión esperada y la desviación estándar de la resistencia a la tensión son de 100 k/pulg^2 y 6 kg/pulg^2, respectivamente. Sea $\overline{X}$ = la resistencia a la tensión promedio de una muestra aleatoria de 40 especímenes tipo A, y sea $\overline{Y}$ = la resistencia a la tensión promedio de una muestra aleatoria de 35 especímenes tipo B.

a. ¿Cuál es la distribución aproximada de $\overline{X}$? ¿De $\overline{Y}$?

b. ¿Cuál es la distribución aproximada de $\overline{X} - \overline{Y}$? Justifique su respuesta.

c. Calcule (aproximadamente) $P(-1 \le \overline{X} - \overline{Y} \le 1)$.

d. Calcule $P(\overline{X} - \overline{Y} \ge 10)$. Si realmente observa que $\overline{X} - \overline{Y} \ge 10$, ¿le cabe duda de que $\mu_1 - \mu_2 = 5$?

74. En un área de suelo arenoso se plantaron 50 árboles pequeños de un cierto tipo, y otros 50 se plantaron en un área de suelo arcilloso. Sea X = el número de árboles plantados en suelo arenoso que sobreviven 1 año y Y = el número de árboles plantados en suelo arcilloso que sobreviven 1 año. Si la probabilidad de que un árbol plantado en suelo arenoso sobreviva 1 año es de .7 y la probabilidad de sobrevivencia de 1 año en suelo arcilloso es de .6, calcule una aproximación a $P(-5 \le X - Y \le 5)$ (no se moleste con la corrección de continuidad).

EJERCICIOS SUPLEMENTARIOS (75-96)

75. Un restaurante sirve tres comidas que cuestan \$12, \$15 y \$20. Para una pareja seleccionada al azar que está comiendo en este restaurante, sea X = el costo de la comida del hombre y Y = el costo de la comida de la mujer. La función de masa de probabilidad conjunta de X y Y se da en la siguiente tabla:

$p(x, y)$		y = 12	15	20
	12	.05	.05	.10
x	15	.05	.10	.35
	20	0	.20	.10

a. Calcule las funciones de masa de probabilidad marginal de X y Y.

b. ¿Cuál es la probabilidad de que las comidas del hombre y la mujer cuesten cuando mucho \$15 cada una?

c. ¿Son X y Y independientes? Justifique su respuesta.

d. ¿Cuál es el costo total esperado de la comida de las dos personas?

e. Suponga que cuando una pareja abre las galletas de la fortuna al final de la comida, encuentran el mensaje: "Recibirá como reembolso la diferencia entre el costo de la comida más cara y la menos cara que eligió". ¿Cuánto espera reembolsar el restaurante?

76. En una estimación de costos, el costo total de un proyecto es la suma de los costos de las tareas componentes. Cada uno de estos costos es una variable aleatoria con una distribución de probabilidad. Se acostumbra obtener información sobre la distribución de costos total sumando las características de las distribuciones de costo de componente individuales; esto se conoce como procedimiento de "despliegue". Por ejemplo, $E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n)$, así que el procedimiento de despliegue es válido para costo medio. Suponga que hay dos tareas componentes y que X_1 y X_2 son variables aleatorias independientes normalmente distribuidas. ¿Es válido el procedimiento de despliegue para el 75º percentil? Es decir, ¿Es el 75º percentil de la distribución de $X_1 + X_2$ el mismo que la suma de los 75º percentiles de las dos distribuciones individuales? Si no, ¿cuál es la relación entre el percentil de la suma y la suma de los percentiles? ¿Con qué percentiles es válido el procedimiento de despliegue en este caso?

77. Una tienda de comida saludable vende dos marcas diferentes de un tipo de grano. Sea X = la cantidad (lb) de la marca A disponible y Y = la cantidad de la marca B disponible. Suponga que la función de densidad de probabilidad conjunta de X y Y es

$$f(x, y) = \begin{cases} kxy & x \geq 0, y \geq 0, 20 \leq x + y \leq 30 \\ 0 & \text{de lo contrario} \end{cases}$$

a. Trace la región de densidad positiva y determine el valor de k.

b. ¿Son X y Y independientes? Responda obteniendo primero la función de densidad de probabilidad marginal de cada variable.

c. Calcule $P(X + Y \leq 25)$.

d. ¿Cuál es la cantidad total esperada de este grano disponible?

e. Calcule $\text{Cov}(X, Y)$ y $\text{Corr}(X, Y)$.

f. ¿Cuál es la varianza de la cantidad total de grano disponible?

78. El artículo "Stochastic Modeling for Pavement Warranty Cost Estimation" (*J. of Constr. Engr. and Mgmnt.*, 2009: 352−359) propone el siguiente modelo para la distribución de Y = tiempo de falla para pavimento. Sean X_1 el momento de un fallo debido a la formación de roderas y sea X_2 el momento de un fallo debido a agrietamiento transversal; estas dos variables aleatorias se suponen independientes. Entonces $Y = \min(X_1, X_2)$. La probabilidad de un fallo debido a cualquiera de estos modos de alteración se supone que es una función creciente del tiempo t. Después de hacer algunas suposiciones de distribución, se obtiene la siguiente forma de la función de distribución acumulativa para cada modo:

$$\Phi\left[(a + bt)/(c + dt + et^2)^{1/2}\right]$$

donde Φ es la función de distribución acumulativa normal estándar. Los valores de los cinco parámetros a, b, c, d y e son −25.49, 1.15, 4.45, −1.78 y .171 para el agrietamiento, y −21.27, .0325, .972, −.00028 y .00022 para las roderas. Determine la probabilidad de falla de los pavimentos en $t = 5$ años y $t = 10$ años.

79. Suponga que para un individuo, la ingesta de calorías en el desayuno es una variable aleatoria con valor esperado de 500 y desviación estándar de 50, la ingesta de calorías en el almuerzo es aleatoria con valor esperado de 900 y desviación estándar de 100, y la ingesta de calorías en la comida es una variable aleatoria con valor esperado de 2000 y desviación estándar de 180. Suponiendo que las ingestas en las diferentes comidas son independientes entre sí, ¿cuál es la probabilidad de que la ingesta de calorías promedio por día durante el siguiente año (365 días) sea cuando mucho de 3500? [*Sugerencia*: sean X_i, Y_i y Z_i las tres ingestas de calorías en el día i. Entonces la ingesta total es $\Sigma(X_i + Y_i + Z_i)$.]

80. El peso medio del equipaje documentado por un pasajero de clase turista seleccionado al azar que vuela entre dos ciudades en cierta aerolínea es de 40 lb y la desviación estándar es de 10 lb. La media y la desviación estándar de un pasajero de clase de negocios son 30 lb y 6 lb, respectivamente.

a. Si hay 12 pasajeros de clase de negocios y 50 de clase turista en un vuelo particular, ¿cuáles son el valor esperado y la desviación estándar del peso total del equipaje?

b. Si los pesos individuales de los equipajes son variables aleatorias independientes normalmente distribuidas, ¿cuál es la probabilidad de que el peso total del equipaje sea cuando mucho de 2500 lb?

81. Se ha visto que si $E(X_1) = E(X_2) = \cdots = E(X_n) = \mu$, entonces $E(X_1 + \cdots + X_n) = n\mu$. En algunas aplicaciones, el número de X_i consideradas no es un número fijo n sino una variable aleatoria N. Por ejemplo, sea N = el número de componentes que son traídos a un taller de reparación en un día particular y sea X_i el tiempo de reparación del componente i-ésimo. Entonces el tiempo de reparación total es $X_1 + X_2 + \cdots + X_N$, la suma de un número *aleatorio* de variables aleatorias. Cuando N es independiente de las X_i se puede demostrar que

$$E(X_1 + \cdots + X_N) = E(N) \cdot \mu$$

a. Si el número esperado de componentes traídos en un día particular es 10 y el tiempo de reparación esperado de un componente seleccionado al azar es de 40 min, ¿cuál es el tiempo de reparación total esperado de componentes entregados en cualquier día particular?

b. Suponga que componentes de un tipo llegan para ser reparados de acuerdo con un proceso de Poisson a razón de 5 por hora. El número esperado de defectos por componente es de 3.5. ¿Cuál es el valor esperado del número total de defectos en componentes traídos a reparación durante un periodo de 4 horas? Asegúrese de indicar cómo su respuesta se deriva del resultado general que se acaba de dar.

82. Suponga que la proporción de votantes rurales en un estado que favorecen a un candidato a gobernador particular es de .45, y que la proporción de votantes suburbanos y urbanos que favorecen al candidato es de .60. Si se obtiene una muestra de 200 votantes rurales y 300 votantes suburbanos y urbanos, ¿cuál es la probabilidad aproximada de que por lo menos 250 de estos votantes favorezcan a este candidato?

83. Sea μ el pH verdadero de un compuesto químico. Se realizará una secuencia de n determinaciones de pH muestrales independientes. Suponga que cada pH muestras es una variable aleato-

ria con valor esperado μ y desviación estándar de .1. ¿Cuántas determinaciones se requieren si se desea que la probabilidad de que el promedio muestral esté dentro de .02 del pH verdadero sea por lo menos de .95? ¿Qué teorema justifica su cálculo de probabilidad?

84. Si la cantidad de refresco que consumo en cualquier día dado es independiente del consumo en cualquier otro día y está normalmente distribuido con $\mu = 13$ oz y $\sigma = 2$, y si en este momento tengo dos paquetes de seis botellas de 16 oz, ¿cuál es la probabilidad de que todavía tenga algo de refresco al cabo de 2 semanas (14 días)?

85. Remítase al ejercicio 58 y suponga que las X_i son independientes entre sí y que cada una tiene una distribución normal. ¿Cuál es la probabilidad de que el volumen total embarcado sea cuando mucho de 100,000 pies^3?

86. Un estudiante tiene una clase que se supone termina a las 9:00 a.m. y otra que se supone comienza a las 9:10 a.m. Suponga que el tiempo real de terminación de la clase de las 9:00 a.m. es una variable aleatoria normalmente distribuida X_1 con media de 9:02 y desviación estándar de 1.5 min y que la hora de inicio de la siguiente clase también es una variable aleatoria normalmente distribuida X_2 con media de 9:10 y desviación estándar de 1 min. Suponga que el tiempo necesario para ir de un salón de clases al otro es una variable aleatoria normalmente distribuida X_3 con media de 6 min y desviación estándar de 1 min. ¿Cuál es la probabilidad de que el estudiante llegue a la segunda clase antes de que comience? (Suponga independencia de X_1, X_2 y X_3, lo cual es razonable si el estudiante no presta atención a la hora de terminación de la primera clase.)

87. **a.** Use la fórmula general de la varianza de una combinación lineal para escribir una expresión para $V(aX + Y)$. Luego sea $a = \sigma_Y/\sigma_X$ y demuestre que $\rho \geq -1$. [*Sugerencia*: la varianza siempre es ≥ 0 y $\text{Cov}(X, Y) = \sigma_X \cdot \sigma_Y \cdot \rho$.]

b. Considerando $V(aX - Y)$, concluya que $\rho \leq 1$.

c. Use el hecho de que $V(W) = 0$ sólo si W es una constante para demostrar que $\rho = 1$ sólo si $Y = aX + b$.

88. Suponga que una calificación oral X y una calificación cuantitativa Y de un individuo seleccionado al azar en un examen de aptitud administrado nacionalmente tienen una función de densidad de probabilidad conjunta

$$f(x, y) = \begin{cases} \frac{2}{5}(2x + 3y) & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{de lo contrario} \end{cases}$$

Se le pide que haga una predicción t de la calificación total $X + Y$ del individuo. El error de predicción es la media del error al cuadrado $E[(X + Y - t)^2]$. ¿Qué valor de t reduce al mínimo el error de predicción?

89. **a.** Sea X_i que tiene una distribución ji cuadrada con parámetro ν_1 (véase la sección 4.4) y sea X_2 independiente de X_1 que tiene una distribución ji cuadrada con parámetro ν_2. Use la técnica del ejemplo 5.21 para demostrar que $X_1 + X_2$ tiene una distribución ji cuadrada con parámetro $\nu_1 + \nu_2$.

b. En el ejercicio 71 del capítulo 4, se le pidió que demostrara que si Z es una variable aleatoria normal estándar, entonces Z^2 tiene una distribución ji cuadrada con $\nu = 1$. Sean $Z_1, Z_2, \ldots, Z_n$ n variables aleatorias normales estándar independientes. ¿Cuál es la distribución de $Z_1^2 + \cdots + Z_n^2$? Justifique su respuesta.

c. Sean $X_1, \ldots, X_n$ una muestra aleatoria de una distribución normal con media μ y varianza σ^2. ¿Cuál es la distribución de la suma $Y = \sum_{i=1}^{n} [(X_i - \mu)/\sigma]^2$? Justifique su respuesta.

90. **a.** Demuestre que $\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$.

b. Sean X_1 y X_2 calificaciones cuantitativas y orales en un examen de aptitud, y sean Y_1 y Y_2 calificaciones correspondientes en otro examen. Si $\text{Cov}(X_1, Y_1) = 5$, $\text{Cov}(X_1, Y_2) = 1$, $\text{Cov}(X_2, Y_1) = 2$ y $\text{Cov}(X_2, Y_2) = 8$, ¿cuál es la covarianza entre las dos calificaciones totales $X_1 + X_2$ y $Y_1 + Y_2$?

91. Se selecciona al azar y se pesa dos veces un espécimen de roca de un área particular. Sea W el peso real y X_1 y X_2 los dos pesos medidos. Entonces $X_1 = W + E_1$ y $X_2 = W + E_2$, donde E_1 y E_2 son los dos errores de medición. Suponga que los E_i son independientes entre sí y de W, y que $V(E_1) = V(E_2) = \sigma_E^2$.

a. Exprese ρ, el coeficiente de correlación entre los dos pesos medidos X_1 y X_2 en función de σ_W^2, la varianza del peso real y σ_X^2, la varianza del peso medido.

b. Calcule ρ cuando $\sigma_W = 1$ kg y $\sigma_E = .01$ kg.

92. Sea A el porcentaje de un constituyente en un espécimen de roca seleccionado al azar y sea B el porcentaje de un segundo constituyente en ese mismo espécimen. Suponga que D y E son errores de medición al determinar los valores de A y B de modo que los valores medidos sean $X = A + D$ y $Y = B + E$, respectivamente. Suponga que los errores de medición son independientes entre sí y de los valores reales.

a. Demuestre que
$\text{Corr}(X, Y) = \text{Corr}(A, B) \cdot \sqrt{\text{Corr}(X_1, X_2)} \cdot \sqrt{\text{Corr}(Y_1, Y_2)}$
donde X_1 y X_2 son mediciones replicadas del valor de A y Y_1 y Y_2 se definen análogamente con respecto a B. ¿Qué efecto tiene la presencia del error de medición en la correlación?

b. ¿Cuál es valor máximo de $\text{Corr}(X, Y)$ cuando $\text{Corr}(X_1, X_2) = .8100$ y $\text{Corr}(Y_1, Y_2) = .9025$? ¿Es esto perturbador?

93. Sean $X_1, \ldots, X_n$ variables aleatorias independientes con valores medios $\mu_1, \ldots, \mu_n$ y varianzas $\sigma_1^2, \ldots, \sigma_n^2$. Considere una función $h(x_1, \ldots, x_n)$ y úsela para definir una nueva variable aleatoria $Y = h(X_1, \ldots, X_n)$. En condiciones un tanto generales en cuanto a la función h, si las σ_i son pequeñas con respecto a las μ_i correspondientes se puede demostrar que $E(Y) \approx h(\mu_1, \ldots, \mu_n)$ y

$$V(Y) \approx \left(\frac{\partial h}{\partial x_1}\right)^2 \cdot \sigma_1^2 + \cdots + \left(\frac{\partial h}{\partial x_n}\right)^2 \cdot \sigma_n^2$$

donde cada derivada parcial se evalúa en $(x_1, \ldots, x_n) = (\mu_1, \ldots, \mu_n)$. Suponga tres resistores con resistencias X_1, X_2, X_3 conectadas en paralelo a través de una batería con voltaje X_4. Luego, según la ley de Ohm, la corriente es

$$Y = X_4\left[\frac{1}{X_1} + \frac{1}{X_2} + \frac{1}{X_3}\right]$$

Sean $\mu_1 = 10$ ohms, $\sigma_1 = 1.0$ ohm, $\mu_2 = 15$ ohms, $\sigma_2 = 1.0$ ohm, $\mu_3 = 20$ ohms, $\sigma_3 = 1.5$ ohms, $\mu_4 = 120$ V, $\sigma_4 = 4.0$ V. Calcule el valor esperado aproximado y la desviación estándar de la corriente (sugerido por "Random Samplings", *CHEMTECH*, 1984: 696–697).

94. Una aproximación más precisa a $E[h(X_1, \ldots, X_n)]$ en el ejercicio 93 es

$$h(\mu_1, \ldots, \mu_n) + \frac{1}{2}\sigma_1^2\left(\frac{\partial^2 h}{\partial x_1^2}\right) + \cdots + \frac{1}{2}\sigma_n^2\left(\frac{\partial^2 h}{\partial x_n^2}\right)$$

Calcule esto con $Y = h(X_1, X_2, X_3, X_4)$ dada en el ejercicio 93 y compárela con el primer término $h(\mu_1, \ldots, \mu_n)$.

95. Sean X y Y variables aleatorias normales estándar independientes y defina una nueva variable aleatoria como $U = .6X + .8Y$.
a. Determine $\text{Corr}(X, U)$.
b. ¿Cómo modificaría U para obtener $\text{Corr}(X, U) = \rho$ con un valor especificado de ρ?

96. Sean $X_1, X_2, \ldots, X_n$ variables aleatorias que denotan n ofertas independientes para un artículo que está a la venta. Supongamos que cada X_i se distribuye uniformemente en el intervalo [100, 200]. Si el vendedor vende al mejor postor, ¿cuánto puede esperar ganar en la venta? [*Sugerencia*: sean $Y = \text{máx}(X_1, X_2, \ldots, X_n)$. En primer lugar encuentre $F_Y(y)$ al notar que $Y \leq y$ si y sólo si cada X_i es $\leq y$. Luego obtenga la función de densidad de probabilidad y $E(Y)$.]

Bibliografía

Devore, Jay y Kenneth Berk, *Modern Mathematical Statistics with Applications*, Thomson-Brooks/Cole, Belmont, CA, 2007. Una exposición un poco más complicada de temas de probabilidad que en el presente libro.

Olkin, Ingram, Cyrus Derman y Leon Gleser, *Probability Models and Applications* (2a. ed.), Macmillan, Nueva York, 1994. Contiene una cuidadosa y amplia exposición de distribuciones conjuntas, reglas de probabilidad y teoremas de límites.

6 Estimación puntual

INTRODUCCIÓN

Dado un parámetro de interés, tal como la media μ o la proporción p de una población, el objetivo de la estimación puntual es utilizar una muestra para calcular un número que representa en cierto sentido una buena suposición del valor verdadero del parámetro. El número resultante se llama *estimación puntual*. En la sección 6.1 se presentan algunos conceptos generales de estimación puntual. En la sección 6.2 se describen e ilustran dos métodos importantes para obtener estimaciones puntuales: el método de momentos y el método de máxima probabilidad.

6.1 Algunos conceptos generales de estimación puntual

El objetivo de la inferencia estadística casi siempre es sacar algún tipo de conclusión sobre uno o más parámetros (características de la población). Para hacer eso un investigador tiene que obtener datos muestrales de cada una de las poblaciones estudiadas. Las conclusiones pueden entonces basarse en los valores calculados de varias cantidades muestrales. Por ejemplo, sea μ (un parámetro) la resistencia a la ruptura promedio verdadera de conexiones alámbricas utilizadas en la unión de obleas semiconductoras. Se podría tomar una muestra aleatoria de $n = 10$ conexiones y determinar la resistencia a la ruptura de cada una y se tendrían las resistencias observadas $x_1, x_2, \ldots, x_{10}$. La resistencia a la ruptura media muestral $\bar{x}$ se utilizaría entonces para sacar una conclusión con respecto al valor de μ. Asimismo, si σ^2 es la varianza de la distribución de la resistencia a la ruptura (varianza de la población, otro parámetro), el valor de la varianza muestral s^2 se utiliza para inferir algo sobre σ^2.

Cuando se discuten los métodos y conceptos generales de inferencia, es conveniente disponer de un símbolo genérico para el parámetro de interés. Se utilizará la letra griega θ para este propósito. El objetivo de la estimación puntual es seleccionar un solo número, con base en los datos muestrales, que represente un valor sensible de θ. Supóngase, por ejemplo, que el parámetro de interés es μ, la vida útil promedio verdadera de baterías de un tipo. Una muestra aleatoria de $n = 3$ baterías podría dar las vidas útiles (horas) observadas $x_1 = 5.0$, $x_2 = 6.4$, $x_3 = 5.9$. El valor calculado de la media muestral de la vida útil es $\bar{x} = 5.77$ y es razonable considerar 5.77 como un valor muy factible de μ, la "mejor suposición" del valor de μ basado en la información muestral disponible.

Supóngase que se desea estimar un parámetro de una sola población (p. ej., μ o σ) con una muestra aleatoria de tamaño n. Recuérdese por el capítulo previo de que antes de que los datos estén disponibles, las observaciones muestrales deben ser consideradas como variables aleatorias $X_1, X_2, \ldots, X_n$. Se deduce que cualquier función de las X_i, es decir, cualquier estadístico, tal como la media muestral $\bar{X}$ o la desviación estándar muestral S también es una variable aleatoria. Lo mismo es cierto si los datos disponibles se componen de más de una muestra. Por ejemplo, se pueden representar las resistencias a la tensión de m especímenes de tipo 1 y de n especímenes de tipo 2 por $X_1, \ldots, X_m$ y $Y_1, \ldots, Y_n$, respectivamente. La diferencia entre las dos medias muestrales de las resistencias es $\bar{X} - \bar{Y}$, el estadístico natural para inferir sobre $\mu_1 - \mu_2$, la diferencia entre las resistencias medias de la población.

DEFINICIÓN

Una **estimación puntual** de un parámetro θ es un número único que puede ser considerado como un valor sensible de θ. Se obtiene una estimación puntual seleccionando un estadístico apropiado y calculando su valor con los datos muestrales dados. El estadístico seleccionado se llama **estimador puntual** de θ.

En el ejemplo de las baterías que se acaba de dar, el estimador utilizado para obtener la estimación puntual de μ fue $\bar{X}$, y la estimación puntual de μ fue 5.77. Si las tres vidas útiles hubieran sido $x_1 = 5.6$, $x_2 = 4.5$ y $x_3 = 6.1$, el uso del estimador $\bar{X}$ habría dado por resultado la estimación $\bar{x} = (5.6 + 4.5 + 6.1)/3 = 5.40$. El símbolo $\hat{\theta}$ ("teta testada") se utiliza comúnmente para denotar tanto la estimación de θ como la estimación puntual que resulta de una muestra dada.* Por tanto, $\hat{\mu} = \bar{X}$ se lee como "el estimador puntual de

* Siguiendo la primera notación, se podría utilizar $\hat{\Theta}$ (una teta mayúscula) para el estimador, pero ésta es difícil de escribir.

μ es la media muestral $\overline{X}$". El enunciado "la estimación puntual de μ es 5.77" se escribe concisamente como $\hat{\mu} = 5.77$. Obsérvese que cuando se escribe $\hat{\theta} = 72.5$, no hay ninguna indicación de cómo se obtuvo esta estimación puntual (qué estadístico se utilizó). Se recomienda reportar tanto el estimador como la estimación resultante.

Ejemplo 6.1 Un fabricante automotriz ha producido un nuevo tipo de parachoques, el que se presume absorbe impactos con menos daño que los parachoques previos. El fabricante lo ha utilizado en una secuencia de 25 choques controlados contra un muro, cada uno a 10 mph, utilizando uno de sus modelos de automóvil compacto. Sea X = el número de choques que no provocaron daños visibles al automóvil. El parámetro que tiene que ser estimado es p = la proporción de todos los choques que no provocaron daños [alternativamente, $p = P$ (ningún daño en un choque)]. Si se observa que X es $x = 15$, el estimador y estimación más razonables son

$$\text{estimador } \hat{p} = \frac{X}{n} \qquad \text{estimación} = \frac{x}{n} = \frac{15}{25} = .60$$ ■

Si por cada parámetro de interés hubiera sólo un estimador puntual razonable, no habría mucho para la estimación puntual. En la mayoría de los problemas, sin embargo, habrá más de un estimador razonable.

Ejemplo 6.2 Reconsidere las 20 observaciones adjuntas de voltaje de ruptura dieléctrica de piezas de resina epóxica introducidas por primera vez en el ejemplo 4.30 (sección 4.6).

24.46	25.61	26.25	26.42	26.66	27.15	27.31	27.54	27.74	27.94
27.98	28.04	28.28	28.49	28.50	28.87	29.11	29.13	29.50	30.88

El patrón en la gráfica de probabilidad normal dado allí es bastante recto, así que ahora se supone que la distribución de voltaje de ruptura es normal con valor medio μ. Como las distribuciones normales son simétricas, μ también es la vida útil mediana de la distribución. Se supone entonces que las observaciones dadas son el resultado de una muestra aleatoria $X_1, X_2, \ldots, X_{20}$ de esta distribución normal. Considere los siguientes estimadores y las estimaciones resultantes de μ:

a. Estimador = $\overline{X}$, estimación = $\bar{x} = \Sigma x_i/n = 555.86/20 = 27.793$

b. Estimador = $\widetilde{X}$, estimación = $\tilde{x} = (27.94 + 27.98)/2 = 27.960$

c. Estimador = $[\text{mín}(X_i) + \text{máx}(X_i)]/2$ = el promedio de las dos vidas útiles extremas, estimación = $[\text{mín}(x_i) + \text{máx}(x_i)]/2 = (24.46 + 30.88)/2 = 27.670$

d. Estimador = $\overline{X}_{\text{rec}(10)}$, la media recortada 10% (desechar el 10% más pequeño y más grande de la muestra y luego promediar),

$$\begin{aligned}\text{estimación} &= \bar{x}_{\text{tr}(10)} \\ &= \frac{555.86 - 24.46 - 25.61 - 29.50 - 30.88}{16} \\ &= 27.838\end{aligned}$$

Cada uno de los estimadores (a)–(d) utiliza una medición diferente del centro de la muestra para estimar μ. ¿Cuál de las estimaciones se acerca más al valor verdadero? No se puede responder esta pregunta sin conocer el valor verdadero. Una pregunta que se puede hacer es: "¿cuál estimador, cuando se utiliza en otras muestras de X_i, tiende a producir estimaciones cercanas al valor verdadero? En breve se considerará este tipo de pregunta. ■

Ejemplo 6.3 El artículo "Is a Normal Distribution the Most Appropriate Statistical Distribution for Volumetric Properties in Asphalt Mixtures?" citado antes en el ejemplo 4.26, informó de las siguientes observaciones sobre X = *vacíos llenos de asfalto* (%) de 52 muestras de un cierto tipo de mezcla caliente de asfalto:

74.33	71.07	73.82	77.42	79.35	82.27	77.75	78.65	77.19
74.69	77.25	74.84	60.90	60.75	74.09	65.36	67.84	69.97
68.83	75.09	62.54	67.47	72.00	66.51	68.21	64.46	64.34
64.93	67.33	66.08	67.31	74.87	69.40	70.83	81.73	82.50
79.87	81.96	79.51	84.12	80.61	79.89	79.70	78.74	77.28
79.97	75.09	74.38	77.67	83.73	80.39	76.90		

Estimemos la varianza σ^2 de la distribución de la población. Un estimador natural es la varianza de la muestra:

$$\hat{\sigma}^2 = S^2 = \frac{\sum(X_i - \bar{X})^2}{n - 1}$$

Minitab dio el siguiente resultado a una petición para mostrar los estadísticos descriptivos:

Variable	Count	Mean	SE Mean	StDev	Variance	Q1	Median	Q3
VFA(B)	52	73.880	0.889	6.413	41.126	67.933	74.855	79.470

Por tanto la estimación puntual de la varianza de la población es

$$\hat{\sigma}^2 = s^2 = \frac{\sum(x_i - \bar{x})^2}{52 - 1} = 41.126$$

[de forma alternativa, la fórmula de cálculo para el numerador de s^2 da

$S_{xx} = \sum x_i^2 - (\sum x_i)^2 / n = 285{,}929.5964 - (3841.78)^2 / 52 = 2097.4124$].

Una estimación puntual de la desviación estándar de la población es entonces $\hat{\sigma} = s = \sqrt{41.126} = 6.413$.

Un estimador alternativo resulta de usar el divisor n en lugar de $n - 1$:

$$\hat{\sigma}^2 = \frac{\sum(X_i - \bar{X})^2}{n}, \quad \text{estimado} = \frac{2097.4124}{52} = 40.335$$

En breve se indicará por qué muchos estadísticos prefieren S^2 en vez de este último estimador.

El citado artículo considera ajustar cuatro distribuciones diferentes a los datos: normal, lognormal, de dos parámetros de Weibull y de tres parámetros de Weibull. Diferentes técnicas se utilizaron para concluir que los dos parámetros de Weibull proporcionan el mejor ajuste (el gráfico de probabilidad normal de los datos muestra alguna desviación de un patrón lineal). De la sección 4.5, la varianza de una variable aleatoria de Weibull es

$$\sigma^2 = \beta^2\{\Gamma(1 + 2/\alpha) - [\Gamma(1 + 1/\alpha)]^2\}$$

donde α y β son los parámetros de forma y escala de la distribución. Los autores del artículo utilizaron el método de máxima verosimilitud (ver sección 6.2) para estimar estos parámetros. Las estimaciones resultantes son $\hat{\alpha} = 11.9731$, $\hat{\beta} = 77.0153$. Una estimación razonable de la varianza de la población ahora se puede obtener al sustituir las estimaciones de los dos parámetros en la expresión para σ^2; el resultado es $\hat{\sigma}^2 = 56.035$. Esta última estimación es obviamente muy diferente de la varianza de la muestra. Su validez depende de que la distribución de la población sea Weibull, mientras que la varianza de la muestra es una manera sensata para estimar σ^2 cuando hay incertidumbre en cuanto a la forma específica de la distribución de la población. ■

En el mejor de todos los mundos posibles, se podría hallar un estimador $\hat{\theta}$ con el cual $\hat{\theta} = \theta$ siempre. Sin embargo, $\hat{\theta}$ es una función de las X_i muestrales, así que es una varia-

ble aleatoria. Con algunas muestras, $\hat{\theta}$ dará un valor más grande que θ, mientras que con otras muestras $\hat{\theta}$ subestimará θ. Si se escribe

$$\hat{\theta} = \theta + \text{error de estimación}$$

entonces un estimador preciso sería uno que produzca errores de estimación pequeños, de manera que los valores estimados estén cerca del valor verdadero.

Una forma sensible de cuantificar la idea de $\hat{\theta}$ cercano a θ es considerar el error al cuadrado $(\hat{\theta} - \theta)^2$. Con algunas muestras, $\hat{\theta}$ se acercará bastante a θ y el error al cuadrado resultante se aproximará a 0. Otras muestras pueden dar valores de $\hat{\theta}$ alejados de θ, correspondientes a errores al cuadrado muy grandes. Una medida general de precisión es *la esperanza o error cuadrático medio* ECM $= E[(\hat{\theta} - \theta)^2]$. Si un primer estimador tiene una ECM más pequeña que un segundo, es natural decir que el primer estimador es el mejor. Sin embargo, el ECM en general dependerá del valor de θ. Lo que a menudo sucede es que un estimador tendrá un ECM más pequeño con algunos valores de θ y un ECM más grande con otros valores. En general no es posible determinar un estimador con el ECM más pequeño.

Una forma de librarse de este dilema es limitar la atención sólo en estimadores que tengan una propiedad deseable específica y luego determinar el mejor estimador en este grupo limitado. Una propiedad popular de esta clase en la comunidad estadística es la *ausencia de sesgo*.

Estimadores insesgados

Supóngase que se tienen dos instrumentos de medición: uno ha sido calibrado con precisión, pero el otro sistemáticamente da lecturas más pequeñas que el valor verdadero que se está midiendo. Cuando cada uno de los instrumentos se utiliza repetidamente en el mismo objeto, debido al error de medición, las mediciones observadas no serán idénticas. Sin embargo, las mediciones producidas por el primer instrumento se distribuirán en torno al valor verdadero de tal modo que en promedio este instrumento mide lo que se propone medir, por lo que se le conoce como instrumento insesgado. El segundo instrumento proporciona observaciones que tienen un componente de error o sesgo sistemático.

DEFINICIÓN

Se dice que un estimador puntual $\hat{\theta}$ es un **estimador insesgado** de θ si $E(\hat{\theta}) = \theta$ para todo valor posible de θ. Si $\hat{\theta}$ es insesgado, la diferencia $E(\hat{\theta}) - \theta$ se conoce como el **sesgo** de $\hat{\theta}$.

Es decir, $\hat{\theta}$ es sesgado si su distribución de probabilidad (es decir, muestreo) siempre está "centrada" en el valor verdadero del parámetro. Supóngase que $\hat{\theta}$ es un estimador insesgado; entonces si $\theta = 100$, la distribución muestral $\hat{\theta}$ está centrada en 100; si $\theta = 27.5$, en ese caso la distribución muestral $\hat{\theta}$ está centrada en 27.5, y así sucesivamente. La figura 6.1 ilustra la distribución de varios estimadores sesgados e insesgados. Obsérvese que "centrada" en este caso significa que el valor esperado, no la mediana, de la distribución de $\hat{\theta}$ es igual a θ.

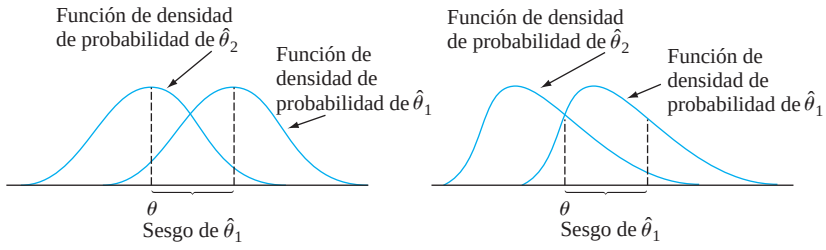


Figura 6.1 Funciones de densidad de probabilidad de un estimador sesgado $\hat{\theta}_1$ y un estimador insesgado $\hat{\theta}_2$, de un parámetro θ

Parece como si fuera necesario conocer el valor de θ (en cuyo caso la estimación es innecesaria) para ver si $\hat{\theta}$ es insesgado. Éste no suele ser el caso, sin embargo, como la ausencia de sesgo es una propiedad general de la distribución muestral del estimador, donde está centrada, lo que por lo general no depende de algún valor de parámetro particular.

En el ejemplo 6.1 se utilizó la proporción muestral X/n como estimador de p, donde X, el número de éxitos muestrales, tenía una distribución binomial con parámetros n y p. Por lo tanto

$$E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{1}{n}\,E(X) = \frac{1}{n}(np) = p$$

PROPOSICIÓN

Cuando X es una variable aleatoria binomial con parámetros n y p, la proporción muestral $\hat{p} = X/n$ es un estimador sesgado de p.

No importa cuál sea el valor verdadero de p, la distribución del estimador $\hat{p}$ estará centrada en el valor verdadero.

Ejemplo 6.4

Suponga que X, el tiempo de reacción a un estímulo, tiene una distribución uniforme en el intervalo desde 0 hasta un límite superior desconocido θ (por tanto la función de densidad de X es de forma rectangular con altura $1/\theta$ en el intervalo $0 \le x \le \theta$). Se desea estimar θ con base en una muestra aleatoria $X_1, X_2, \ldots, X_n$ de los tiempos de reacción. Como θ es el tiempo más grande posible en toda la población de tiempos de reacción, considere como un primer estimador el tiempo de reacción muestral más grande $\hat{\theta}_1 = \text{máx}(X_1, \ldots, X_n)$. Si $n = 5$ y $x_1 = 4.2$, $x_2 = 1.7$, $x_3 = 2.4$, $x_4 = 3.9$, y $x_5 = 1.3$, la estimación puntual de θ es $\hat{\theta}_1 = \text{máx}(4.2, 1.7, 2.4, 3.9, 1.3) = 4.2$.

La ausencia de sesgo implica que algunas muestras darán estimaciones que exceden θ y otras que darán estimaciones más pequeñas que θ, de lo contrario θ posiblemente no podría ser el centro (punto de equilibrio) de la distribución de $\hat{\theta}_1$. Sin embargo, el estimador propuesto nunca sobrestimará θ (el valor muestral más grande no puede exceder el valor de la población más grande) y subestimará θ a menos que el valor muestral más grande sea igual a θ. Este argumento intuitivo demuestra que $\hat{\theta}_1$ es un estimador insesgado. Más precisamente, se puede demostrar (véase el ejercicio 32) que

$$E(\hat{\theta}_1) = \frac{n}{n+1} \cdot \theta < \theta \qquad \left(\text{como } \frac{n}{n+1} < 1\right)$$

El sesgo de $\hat{\theta}_1$ está dado por $n\theta/(n+1) - \theta = -\theta/(n+1)$, el cual tiende a cero a medida que n se hace grande.

Es fácil modificar $\hat{\theta}_1$ para obtener un estimador insesgado de θ. Considere el estimador

$$\hat{\theta}_2 = \frac{n+1}{n} \cdot \text{máx}(X_1, \ldots, X_n)$$

Utilizando este estimador en los datos se obtiene la estimación $(6/5)(4.2) = 5.04$. El hecho de que $(n+1)/n > 1$ implica que $\hat{\theta}_2$ sobrestimará θ para algunas muestras y la subestimará en otras. El valor medio de este estimador es

$$E(\hat{\theta}_2) = E\left[\frac{n+1}{n}\,\text{máx}(X_1, \ldots, X_n)\right] = \frac{n+1}{n} \cdot E\left[\text{máx}(X_1, \ldots, X_n)\right]$$
$$= \frac{n+1}{n} \cdot \frac{n}{n+1}\theta = \theta$$

Si $\hat{\theta}_2$ se utiliza repetidamente en diferentes muestras para estimar θ, algunas estimaciones serán demasiado grandes y otras demasiado pequeñas, pero a la larga no habrá ninguna tendencia simétrica a subestimar o sobrestimar θ. ■

Principio de estimación insesgada

Cuando se elige entre varios estimadores diferentes de θ, se elige uno insesgado.

De acuerdo con este principio, el estimador insesgado $\hat{\theta}_2$ en el ejemplo 6.4 deberá ser preferido al estimador sesgado $\hat{\theta}_1$. Considérese ahora el problema de estimar σ^2.

PROPOSICIÓN

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución con media μ y varianza σ^2. Entonces el estimador

$$\hat{\sigma}^2 = S^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$$

es un estimador insesgado de σ^2.

Demostración Para cualquier variable aleatoria Y, $V(Y) = E(Y^2) - [E(Y)]^2$, por lo tanto $E(Y^2) = V(Y) + [E(Y)]^2$. Aplicando esto a

$$S^2 = \frac{1}{n-1}\left[\sum X_i^2 - \frac{(\sum X_i)^2}{n}\right]$$

se obtiene

$$E(S^2) = \frac{1}{n-1}\left\{\sum E(X_i^2) - \frac{1}{n}E[(\sum X_i)^2]\right\}$$
$$= \frac{1}{n-1}\left\{\sum(\sigma^2 + \mu^2) - \frac{1}{n}\{V(\sum X_i) + [E(\sum X_i)]^2\}\right\}$$
$$= \frac{1}{n-1}\left\{n\sigma^2 + n\mu^2 - \frac{1}{n}n\sigma^2 - \frac{1}{n}(n\mu)^2\right\}$$
$$= \frac{1}{n-1}\{n\sigma^2 - \sigma^2\} = \sigma^2 \text{ (como se desea)}$$ ■

El estimador que utiliza el divisor n se expresa como $(n-1)S^2/n$, por lo tanto

$$E\left[\frac{(n-1)S^2}{n}\right] = \frac{n-1}{n}E(S^2) = \frac{n-1}{n}\sigma^2$$

Este estimador es por consiguiente sesgado. El sesgo es $(n-1)\sigma^2/n - \sigma^2 = -\sigma^2/n$. Como el sesgo es negativo, el estimador con divisor n tiende a subestimar σ^2 y por eso muchos estadísticos prefieren el divisor $n-1$ (aunque cuando n es grande, el sesgo es pequeño y hay poca diferencia entre los dos).

Lamentablemente, el hecho de que S^2 sea insesgado para la estimación de σ^2 no implica que S sea insesgado para la estimación de σ. Sacar la raíz cuadrada estropea la propiedad de ausencia de sesgo (el valor esperado de la raíz cuadrada no es la raíz cuadrada

del valor esperado). Afortunadamente, el sesgo de S es pequeño a menos que n sea muy pequeño. Hay otras buenas razones para utilizar S como estimador, especialmente cuando la distribución de la población es normal. Esto se volverá más aparente cuando se discutan los intervalos de confianza y la prueba de hipótesis en los siguientes capítulos.

En el ejemplo 6.2 se propusieron varios estimadores diferentes de la media μ de una distribución normal. Si hubiera un estimador insesgado único para μ, el problema de estimación se resolvería utilizando dicho estimador. Desafortunadamente, este no es el caso.

PROPOSICIÓN

Si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria tomada de una distribución con media μ, entonces $\overline{X}$ es un estimador sesgado de μ. Si además la distribución es continua y simétrica, entonces $\widetilde{X}$ y cualquier media recortada también son estimadores insesgados de μ.

El hecho de que $\overline{X}$ sea insesgado es simplemente un replanteamiento de una de las reglas de valor esperado: $E(\overline{X}) = \mu$ para cada valor posible de μ (para distribuciones discretas y continuas). La ausencia de sesgo de los demás estimadores es más difícil de verificar.

El siguiente ejemplo introduce otra situación en la cual existen varios estimadores insesgados para un parámetro particular.

Ejemplo 6.5 En ciertas circunstancias contaminantes orgánicos se adhieren con facilidad a las superficies de obleas y deterioran los dispositivos de fabricación de semiconductores. El artículo "Ceramic Chemical Filter for Removal of Organic Contaminants" (*J. of the Institute of Environmental Sciences and Technology*, 2003: 59–65) discutió una alternativa recientemente desarrollada de filtros de carbón convencionales para eliminar contaminación molecular transportada por el aire en aplicaciones de cuartos limpios. Un aspecto de la investigación del desempeño de filtros implicó estudiar cómo se relaciona la concentración de contaminantes en el aire con la concentración en la superficie de una oblea después de una exposición prolongada. Considere los siguientes datos representativos de x = concentración de DBP en aire y y = concentración de DBP en la superficie de obleas luego de 4 horas de exposición (ambas en $\mu g/m^3$, donde DBP = ftalato de dibutilo).

Obs. i:	1	2	3	4	5	6
x:	.8	1.3	1.5	3.0	11.6	26.6
y:	.6	1.1	4.5	3.5	14.4	29.1

Los autores comentan que la "adhesión de DBP en la superficie de obleas fue aproximadamente proporcional a la concentración de DBP en el aire". La figura 6.2 muestra una gráfica de y contra x, es decir, de los pares (x, y).

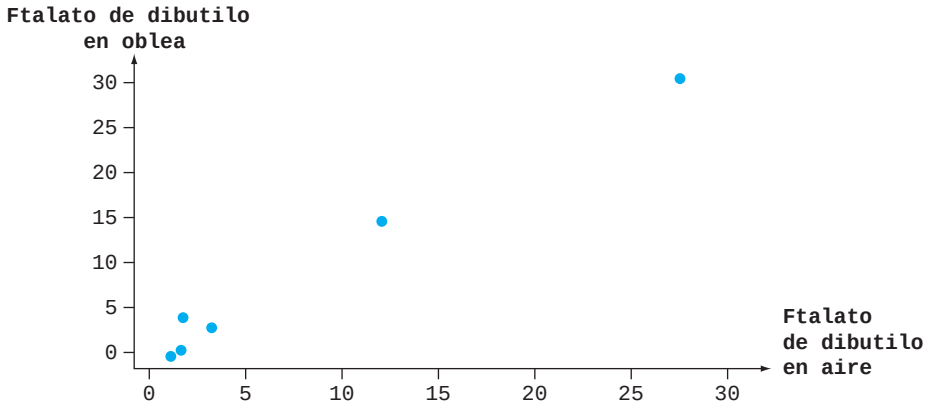


Figura 6.2 Gráfica de los datos de ftalato de dibutilo del ejemplo 6.5

Si y fuera exactamente proporcional a x, se tendría $y = \beta x$ para algún valor β, lo cual expresa que los puntos (x, y) en la gráfica quedarían exactamente sobre una línea recta con pendiente β que pasa por (0, 0). Pero esto es sólo aproximadamente el caso. Así que a continuación se supone que para cualquier x fija, la concentración de DBP en las obleas es una variable aleatoria Y con valor medio βx. Es decir, se postula que el valor *medio* de Y está relacionado con x por una línea que pasa por (0, 0) pero que el valor observado de Y en general se desviará de esta línea (esto se conoce en la literatura estadística como "regresión a través del origen").

Ahora se desea estimar el parámetro de la pendiente β. Considere los siguientes tres estimadores:

$$\#1\colon \hat{\beta} = \frac{1}{n}\sum \frac{Y_i}{x_i} \qquad \#2\colon \hat{\beta} = \frac{\sum Y_i}{\sum x_i} \qquad \#3\colon \hat{\beta} = \frac{\sum x_i Y_i}{\sum x_i^2}$$

Las estimaciones resultantes basadas en los datos dados son 1.3497, 1.1875 y 1.1222, respectivamente. Así que de manera definitiva la estimación depende de qué estimador se utilice. Si uno de estos tres estimadores fuera insesgado y los otros dos sesgados, habría un buen motivo para utilizar el insesgado. Pero los tres son insesgados; el argumento se apoya en el hecho de que cada uno es una función lineal de las Y_i (aquí se supone que las x_i son fijas, no aleatorias):

$$E\left(\frac{1}{n}\sum \frac{Y_i}{x_i}\right) = \frac{1}{n}\sum \frac{E(Y_i)}{x_i} = \frac{1}{n}\sum \frac{\beta x_i}{x_i} = \frac{1}{n}\sum \beta = \frac{n\beta}{n} = \beta$$

$$E\left(\frac{\sum Y_i}{\sum x_i}\right) = \frac{1}{\sum x_i} E\left(\sum Y_i\right) = \frac{1}{\sum x_i}\left(\sum \beta x_i\right) = \frac{1}{\sum x_i}\beta\left(\sum x_i\right) = \beta$$

$$E\left(\frac{\sum x_i Y_i}{\sum x_i^2}\right) = \frac{1}{\sum x_i^2} E\left(\sum x_i Y_i\right) = \frac{1}{\sum x_i^2}\left(\sum x_i \beta x_i\right) = \frac{1}{\sum x_i^2}\beta\left(\sum x_i^2\right) = \beta$$

■

Tanto en el ejemplo anterior como en la situación que implica estimar una media de población normal, el principio de ausencia de sesgo (preferir un estimador insesgado a uno sesgado) no puede ser invocado para seleccionar un estimador. Lo que ahora se requiere es un criterio para elegir entre estimadores sesgados.

Estimadores con varianza mínima

Supóngase que $\hat{\theta}_1$ y $\hat{\theta}_2$ son dos estimadores de θ insesgados. Entonces, aunque la distribución de cada estimador esté centrada en el valor verdadero de θ, las dispersiones de las distribuciones en torno al valor verdadero pueden ser diferentes.

> **Principio de estimación insesgada con varianza mínima**
>
> Entre todos los estimadores de θ insesgados, se selecciona el de varianza mínima. El $\hat{\theta}$ resultante se llama **estimador insesgado con varianza mínima (EIVM)** de θ.

La figura 6.3 ilustra las funciones de densidad de probabilidad de los dos estimadores insesgados, donde $\hat{\theta}_1$ tiene una varianza más pequeña que $\hat{\theta}_2$. Entonces es más probable que $\hat{\theta}_1$ produzca una estimación próxima al valor verdadero θ que $\hat{\theta}_2$. El estimador insesgado con varianza mínima es, en cierto sentido, el que tiene más probabilidades entre todos los estimadores insesgados de producir una estimación cercana al verdadero θ.

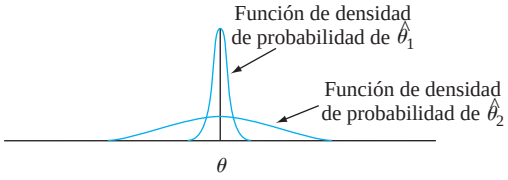


Figura 6.3 Gráficas de las funciones de densidad de probabilidad de dos estimadores insesgados diferentes

En el ejemplo 6.5, supóngase que cada Y_i está normalmente distribuida con media βx_i y varianza σ^2 (la suposición de varianza constante). Entonces se puede demostrar que el tercer estimador $\hat{\beta} = \sum x_i Y_i / \sum x_i^2$ no sólo tiene una varianza más pequeña que cualquiera de los otros dos estimadores insesgados, sino que de hecho es el estimador insesgado con varianza mínima; tiene una varianza más pequeña que *cualquier* otro estimador insesgado de β.

Ejemplo 6.6 En el ejemplo 6.4 se argumentó que cuando $X_1, \ldots, X_n$ es una muestra aleatoria tomada de una distribución uniforme en el intervalo $[0, \theta]$, el estimador

$$\hat{\theta}_1 = \frac{n+1}{n} \cdot \text{máx}(X_1, \ldots, X_n)$$

es insesgado para θ (previamente este estimador se denotó por $\hat{\theta}_2$). Éste no es el único estimador insesgado de θ . El valor esperado de una variable aleatoria uniformemente distribuida es simplemente el punto medio del intervalo de densidad positiva, por lo tanto $E(X_i) = \theta/2$. Esto implica que $E(\overline{X}) = \theta/2$, a partir de la cual $E(2\overline{X}) = \theta$. Es decir, el estimador $\hat{\theta}_2 = 2\overline{X}$ es insesgado para θ.

Si X está uniformemente distribuida en el intervalo de A a B, entonces $V(X) = \sigma^2 = (B - A)^2/12$. Así pues, en esta situación, $V(X_i) = \theta^2/12$, $V(\overline{X}) = \sigma^2/n = \theta^2/(12n)$ y $V(\hat{\theta}_2) = V(2\overline{X}) = 4V(\overline{X}) = \theta^2/(3n)$. Se pueden utilizar los resultados del ejercicio 32 para demostrar que $V(\hat{\theta}_1) = \theta^2/[n(n+2)]$. El estimador $\hat{\theta}_1$ tiene una varianza más pequeña que $\hat{\theta}_2$ si $3n < n(n+2)$; es decir, si $0 < n^2 - n = n(n-1)$. En tanto $n > 1$, $V(\hat{\theta}_1) < V(\hat{\theta}_2)$, así que $\hat{\theta}_1$ es mejor estimador que $\hat{\theta}_2$. Se pueden utilizar métodos más avanzados para demostrar que $\hat{\theta}_1$ es el estimador insesgado con varianza mínima de θ; cualquier otro estimador insesgado de θ tiene una varianza que excede a $\theta^2/[n(n+2)]$. ■

Uno de los triunfos de la estadística matemática ha sido el desarrollo de una metodología para identificar el estimador insesgado con varianza mínima en una amplia variedad de situaciones. El resultado más importante de este tipo para nuestros propósitos tiene que ver con la estimación de la media μ de una distribución normal.

TEOREMA

Sea $X_1, \ldots, X_n$ una muestra aleatoria tomada de una distribución normal con parámetros μ y σ. Entonces el estimador $\hat{\mu} = \overline{X}$ es el estimador insesgado con varianza mínima para μ.

Siempre que exista la seguridad de que la población que se está muestreando es normal, el teorema dice que $\bar{x}$ debe usarse para estimar μ. Entonces, en el ejemplo 6.2 la estimación sería $\bar{x} = 27.793$.

En algunas situaciones es posible obtener un estimador con sesgo pequeño que se preferiría al mejor estimador insesgado. Esto se ilustra en la figura 6.4. Sin embargo, los estimadores insesgados con varianza mínima a menudo son más fáciles de obtener que el tipo de estimador sesgado cuya distribución se ilustra.

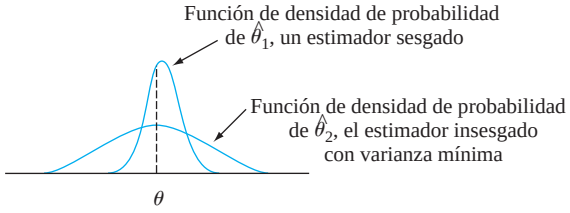


Figura 6.4 Un estimador sesgado que es preferible al estimador insesgado con varianza mínima

Algunas complicaciones

El último teorema no dice que al estimar la media μ de una población, se deberá utilizar el estimador $\overline{X}$ independientemente de la distribución que se está muestreando.

Ejemplo 6.7 Suponga que se desea estimar la conductividad térmica μ de un material. Con técnicas de medición estándar, se obtendrá una muestra aleatoria $X_1, \ldots, X_n$ de n mediciones de conductividad térmica. Suponga que la distribución de la población es un miembro de una de las siguientes tres familias:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x-\mu)^2/(2\sigma^2)} \qquad -\infty < x < \infty \tag{6.1}$$

$$f(x) = \frac{1}{\pi[1 + (x - \mu)^2]} \qquad -\infty < x < \infty \tag{6.2}$$

$$f(x) = \begin{cases} \dfrac{1}{2c} & -c \le x - \mu \le c \\ 0 & \text{de lo contrario} \end{cases} \tag{6.3}$$

La función de densidad de probabilidad (6.1) es la distribución normal, (6.2) se llama distribución de Cauchy y (6.3) es una distribución uniforme. Las tres distribuciones son simétricas con respecto a μ y de hecho la distribución de Cauchy tiene forma de campana pero con colas mucho más gruesas (más probabilidad hacia fuera) que la curva normal. La distribución uniforme no tiene colas. Los cuatro estimadores de μ considerados con anterioridad son $\overline{X}$, $\tilde{X}$, $\overline{X}_e$ (el promedio de las dos observaciones extremas) y $\overline{X}_{\text{rec}(10)}$, una media recortada.

La muy importante moraleja en este caso es que el mejor estimador de μ depende crucialmente de qué distribución está siendo muestreada. En particular,

1. Si la muestra aleatoria proviene de una distribución normal, en ese caso $\overline{X}$ es el mejor de los cuatro estimadores, puesto que tiene una varianza mínima entre todos los estimadores insesgados.
2. Si la muestra aleatoria proviene de una distribución de Cauchy, entonces $\overline{X}$ y $\overline{X}_e$ son estimadores terribles de μ, en tanto que $\tilde{X}$ es bastante bueno (el estimador insesgado con varianza mínima no es conocido); $\overline{X}$ es malo porque es muy sensible a las observaciones subyacentes y las colas gruesas de la distribución de Cauchy hacen que sea improbable que aparezcan tales observaciones en cualquier muestra.
3. Si la distribución subyacente es uniforme, el mejor estimador es $\overline{X}_e$; este estimador está influido en gran medida por las observaciones subyacentes, pero la carencia de colas hace que tales observaciones sean imposibles.
4. *En ninguna de estas tres situaciones es mejor la media recortada pero funciona razonablemente bien en las tres*. Es decir, $\overline{X}_{\text{rec}(10)}$ no sufre demasiado en comparación con el mejor procedimiento en cualquiera de las tres situaciones. ■

Más generalmente, investigaciones recientes en estadística han establecido que cuando se estima un punto de simetría μ de una distribución de probabilidad continua, una media recortada con proporción de recorte de 10% o 20% (para cada extremo de la mues-

tra) produce estimaciones razonablemente comportadas dentro de un rango muy amplio de posibles modelos. Por esta razón, se dice que una media recortada con poco porcentaje de recorte es un **estimador robusto**.

En algunas situaciones, la selección no es entre dos estimadores diferentes construidos con la misma muestra, sino entre estimadores basados en dos experimentos distintos.

Ejemplo 6.8 Suponga que cierto tipo de componente tiene una distribución de vida útil exponencial con parámetro λ de modo que la vida útil esperada es $\mu = 1/\lambda$. Se selecciona una muestra de n de esos componentes y cada uno es puesto en operación. Si el experimento continúa hasta que todas las n vidas útiles, $X_1, \ldots, X_n$ han sido observadas, en ese caso $\overline{X}$ es un estimador insesgado de μ.

En algunos experimentos, sin embargo, los componentes se dejan en operación sólo hasta el tiempo de la falla r-ésima, donde $r < n$. Este procedimiento se conoce como **censura.** Sea Y_1 el tiempo de la primera falla (la vida útil mínima entre los n componentes) y Y_2 el tiempo en el cual ocurre la segunda falla (la segunda vida útil más pequeña), y así sucesivamente. Como el experimento termina en el tiempo Y_r, la vida útil acumulada al final es

$$T_r = \sum_{i=1}^{r} Y_i + (n - r)Y_r$$

A continuación se demuestra que $\hat{\mu} = T_r/r$ es un estimador insesgado de μ. Para hacerlo, se requieren dos propiedades de las variables exponenciales:

1. La propiedad de no memoria (véase la sección 4.4), la cual dice que en cualquier punto de tiempo, la vida útil restante tiene la misma distribución exponencial que la vida útil original.
2. Si $X_1, \ldots, X_k$ son independientes, cada mín$(X_1, \ldots, X_k)$ exponencialmente distribuida con parámetro λ, es exponencial con parámetro $k\lambda$ y su valor esperado es $1/(k\lambda)$.

Como los n componentes duran hasta Y_1, $n - 1$ duran una cantidad de tiempo $Y_2 - Y_1$ adicional, $n - 2$, duran una cantidad de tiempo $Y_3 - Y_2$ adicional, y así sucesivamente, otra expresión para T_r es

$$\begin{aligned} T_r = {} & nY_1 + (n-1)(Y_2 - Y_1) + (n-2)(Y_3 - Y_2) + \cdots \\ & + (n - r + 1)(Y_r - Y_{r-1}) \end{aligned}$$

Pero Y_1 es la mínima de n variables exponenciales, por tanto $E(Y_1) = 1/(n\lambda)$. Asimismo, $Y_2 - Y_1$ es la más pequeña de las $n - 1$ vidas útiles restantes, cada una exponencial con parámetro λ (por la propiedad de no memoria), así que $E(Y_2 - Y_1) = 1/[(n-1)\lambda]$. Continuando, $E(Y_{i+1} - Y_i) = 1/[(n-i)\lambda]$ así que

$$\begin{aligned} E(T_r) &= nE(Y_1) + (n-1)E(Y_2 - Y_1) + \cdots + (n - r + 1)E(Y_r - Y_{r-1}) \\ &= n \cdot \frac{1}{n\lambda} + (n-1) \cdot \frac{1}{(n-1)\lambda} + \cdots + (n - r + 1) \cdot \frac{1}{(n - r + 1)\lambda} \\ &= \frac{r}{\lambda} \end{aligned}$$

Por consiguiente, $E(T_r/r) = (1/r)E(T_r) = (1/r) \cdot (r/\lambda) = 1/\lambda = \mu$ como se proponía.

Como un ejemplo, supóngase que se prueban 20 componentes y $r = 10$. Entonces si los primeros diez tiempos de falla son 11, 15, 29, 33, 35, 40, 47, 55, 58 y 72, la estimación de μ es

$$\hat{\mu} = \frac{11 + 15 + \cdots + 72 + (10)(72)}{10} = 111.5$$

La ventaja del experimento con censura es que termina más rápido que el experimento sin censura. Sin embargo, se puede demostrar que $V(T_r/r) = 1/(\lambda^2 r)$, la cual es más grande que $1/(\lambda^2 n)$, la varianza de $\overline{X}$ en el experimento sin censura. ■

Reporte de una estimación puntual: el error estándar

Además de reportar el valor de una estimación puntual, se debe dar alguna indicación de su precisión. La medición usual de precisión es el error estándar del estimador usado.

DEFINICIÓN

El **error estándar** de un estimador $\hat{\theta}$ es su desviación estándar $\sigma_{\hat{\theta}} = \sqrt{V(\hat{\theta})}$. Éste es la magnitud de una desviación típica o representativa entre una estimación y el valor de θ. Si el error estándar implica parámetros desconocidos cuyos valores pueden ser estimados, la sustitución de estas estimaciones en $\sigma_{\hat{\theta}}$ da el **error estándar estimado** (desviación estándar estimada) del estimador. El error estándar estimado puede ser denotado por $\hat{\sigma}_{\hat{\theta}}$ (el $\hat{}$ sobre σ recalca que $\sigma_{\hat{\theta}}$ está siendo estimada) o por $s_{\hat{\theta}}$.

Ejemplo 6.9 (Continuación del ejemplo 6.2)

Suponiendo que el voltaje de ruptura está normalmente distribuido, $\hat{\mu} = \overline{X}$ es la mejor estimación de μ. Si se sabe que el valor de σ es 1.5, el error estándar de $\overline{X}$ es $\sigma_{\overline{X}} = \sigma/\sqrt{n} = 1.5/\sqrt{20} = .335$. Si, como casi siempre es el caso, el valor de σ es desconocido, la estimación $\hat{\sigma} = s = 1.462$ se sustituye en $\sigma_{\overline{X}}$ para obtener el error estándar estimado $\hat{\sigma}_{\overline{X}} = s_{\overline{X}} = s/\sqrt{n} = 1.462/\sqrt{20} = .327$. ■

Ejemplo 6.10 (Continuación del ejemplo 6.1)

El error estándar de $\hat{p} = X/n$ es

$$\sigma_{\hat{p}} = \sqrt{V(X/n)} = \sqrt{\frac{V(X)}{n^2}} = \sqrt{\frac{npq}{n^2}} = \sqrt{\frac{pq}{n}}$$

Como p y $q = 1 - p$ son desconocidas (o bien ¿por qué estimarlas?), se sustituyen $\hat{p} = x/n$ y $\hat{q} = 1 - x/n$ en $\sigma_{\hat{p}}$, para obtener el error estándar estimado $\hat{\sigma}_{\hat{p}} = \sqrt{\hat{p}\hat{q}/n} = \sqrt{(.6)(.4)/25} = .098$. Alternativamente, como el valor más grande de pq se obtiene cuando $p = q = .5$, un límite superior en el error estándar es $\sqrt{1/(4n)} = .10$. ■

Cuando la distribución del estimador puntual $\hat{\theta}$ es normal de modo aproximado, lo que a menudo será el caso cuando n es grande, entonces se puede estar confiado de manera razonable en que el valor verdadero de θ queda dentro de aproximadamente 2 errores estándar (desviaciones estándar) de $\hat{\theta}$. De este modo, si una muestra de $n = 36$ vidas útiles de componentes da $\hat{\mu} = \bar{x} = 28.50$ y $s = 3.60$, por consiguiente $s/\sqrt{n} = .60$, dentro de dos errores estándar estimados, $\hat{\mu}$ se traslada al intervalo $28.50 \pm (2)(.60) = (27.30, 29.70)$.

Si $\hat{\theta}$ no es necesariamente normal en forma aproximada pero es insesgado, entonces se puede demostrar que la estimación se desviará de θ hasta por 4 errores estándar cuando mucho 6% del tiempo. Se esperaría entonces que el valor verdadero quede dentro de 4 errores estándar de $\hat{\theta}$ (y ésta es una afirmación muy conservadora, puesto que se aplica a *cualquier* $\hat{\theta}$ insesgado). Resumiendo, el error estándar indica de forma aproximada a qué distancia de $\hat{\theta}$ se puede esperar que quede el valor verdadero de θ.

La forma del estimador de $\hat{\theta}$ puede ser suficientemente complicada de modo que la teoría estadística estándar no pueda ser aplicada para obtener una expresión para $\sigma_{\hat{\theta}}$. Esto es cierto, por ejemplo, en el caso $\theta = \sigma$, $\hat{\theta} = S$; la desviación estándar del estadístico S, σ_S, en general no puede ser determinada. No hace mucho se introdujo un método de computadora intensivo llamado *bootstrap* para abordar este problema. Supóngase que la función de densidad de probabilidad de la población es $f(x; \theta)$, un miembro de una familia paramétrica particular, y que los datos $x_1, x_2, \ldots, x_n$ dan $\hat{\theta} = 21.7$. Ahora se utiliza la computadora para obtener "muestras bootstrap" tomadas de la función de densidad de probabilidad $f(x; 21.7)$, y por cada muestra se calcula una "estimación bootstrap" $\hat{\theta}^*$:

Primera muestra "bootstrap" : $x_1^*, x_2^*, \ldots, x_n^*$; estimación $= \hat{\theta}_1^*$
Segunda muestra "bootstrap": $x_1^*, x_2^*, \ldots, x_n^*$; estimación $= \hat{\theta}_2^*$
$\vdots$
B-ésima muestra bootstrap: $x_1^*, x_2^*, \ldots, x_n^*$; estimación $= \hat{\theta}_B^*$

A menudo se utiliza $B = 100$ o 200. Ahora sea $\bar{\theta}^* = \Sigma\hat{\theta}_i^*/B$, la media muestral de las estimaciones "bootstrap". La **estimación bootstrap** del error estándar de $\hat{\theta}$ ahora es simplemente la desviación estándar muestral de los $\hat{\theta}_i^*$:

$$S_{\hat{\theta}} = \sqrt{\frac{1}{B-1}\Sigma(\hat{\theta}_i^* - \bar{\theta}^*)^2}$$

(En la literatura de bootstrap, a menudo se utiliza B en lugar de $B - 1$; con valores típicos de B, casi siempre hay poca diferencia entre las estimaciones resultantes.)

Ejemplo 6.11 Un modelo teórico sugiere que X, el tiempo para la ruptura de un fluido aislante entre electrodos a un voltaje particular, tiene $f(x; \lambda) = \lambda e^{-\lambda x}$, una distribución exponencial. Una muestra aleatoria de $n = 10$ tiempos de ruptura (minutos) da los datos siguientes:

41.53 18.73 2.99 30.34 12.33 117.52 73.02 223.63 4.00 26.78

Como $E(X) = 1/\lambda$, $E(\bar{X}) = 1/\lambda$, una estimación razonable de λ es $\hat{\lambda} = 1/\bar{x} = 1/55.087 = .018153$. Se utilizaría entonces un programa de computadora estadístico para obtener $B = 100$ muestras bootstrap, cada una de tamaño 10, provenientes de $f(x; .018153)$. La primera muestra fue 41.00, 109.70, 16.78, 6.31, 6.76, 5.62, 60.96, 78.81, 192.25, 27.61, con la cual $\Sigma x_i^* = 545.8$ y $\hat{\lambda}_1^* = 1/54.58 = .01832$. El promedio de las 100 estimaciones bootstrap es $\bar{\lambda}^* = .02153$, y la desviación estándar muestral de estas 100 estimaciones es $s_{\hat{\lambda}} = .0091$, la estimación bootstrap del error estándar de $\hat{\lambda}$. Un histograma de los $100\hat{\lambda}_i^*$ resultó un tanto positivamente asimétrico, lo que sugiere que la distribución muestral de $\hat{\lambda}$ también tiene esta propiedad. ■

En ocasiones un investigador desea estimar una característica poblacional sin suponer que la distribución de la población pertenece a una familia paramétrica particular. Una instancia de esto ocurrió en el ejemplo 6.7, cuando una media recortada 10% fue propuesta para estimar el centro θ de una distribución de población simétrica. Los datos del ejemplo 6.2 dieron $\hat{\theta} = \bar{x}_{\text{rec}(10)} = 27.838$, pero ahora no hay ninguna $f(x; \theta)$ supuesta, por consiguiente, ¿cómo se puede obtener una muestra bootstrap? La respuesta es considerar que la muestra constituye la población (las $n = 20$ observaciones en el ejemplo 6.2) y tome B muestras diferentes, cada una de tamaño n, *con* reemplazo, de esta población. El libro de Bradley Efron y Robert Tibshirani o el de John Rice incluidos en la bibliografía del capítulo proporcionan más información.

EJERCICIOS Sección 6.1 (1–19)

1. Los datos adjuntos sobre resistencia a la flexión (MPa) de vigas de concreto de un tipo se introdujeron en el ejemplo 1.2.

5.9	7.2	7.3	6.3	8.1	6.8	7.0
7.6	6.8	6.5	7.0	6.3	7.9	9.0
8.2	8.7	7.8	9.7	7.4	7.7	9.7
7.8	7.7	11.6	11.3	11.8	10.7	

a. Calcule una estimación puntual del valor medio de resistencia de la población conceptual de todas las vigas fabricadas de esta manera, y diga qué estimador utilizó. [*Sugerencia*: $\Sigma x_i = 219.8$.]

b. Calcule una estimación puntual del valor de resistencia que separa el 50% más débil de dichas vigas del 50% más resistente, y diga qué estimador utilizó.

c. Calcule e interprete una estimación puntual de la desviación estándar de la población σ. ¿Qué estimador utilizó? [*Sugerencia*: $\sum x_i^2 = 1860.94$.]

d. Calcule una estimación puntual de la proporción de las vigas cuya resistencia a la flexión exceda de 10 MPa. [*Sugerencia*: considere una observación como "éxito" si excede de 10.]

e. Calcule una estimación puntual del coeficiente de variación de la población σ/μ y diga qué estimador utilizó.

2. Una muestra de 20 estudiantes que recientemente tomaron un curso de estadística elemental arrojó la siguiente información sobre la marca de calculadora que poseían. (T = Texas Instruments, H = Hewlett Packard, C = Casio, S = Sharp):

T T H T C T T S C H
S S T H C T T T H T

a. Estime la verdadera proporción de los estudiantes que poseen una calculadora Texas Instruments.

b. De los 10 estudiantes que poseían una calculadora TI, 4 tenían calculadoras graficadoras. Estime la proporción de estudiantes que no poseen una calculadora graficadora TI.

3. Considere la siguiente muestra de observaciones sobre el espesor de recubrimiento de pintura de baja viscosidad ("Achieving a Target Value for a Manufacturing Process: A Case Study", *J. of Quality Technology*, 1992: 22–26):

.83 .88 .88 1.04 1.09 1.12 1.29 1.31
1.48 1.49 1.59 1.62 1.65 1.71 1.76 1.83

Suponga que la distribución del espesor de recubrimiento es normal (una gráfica de probabilidad normal soporta fuertemente esta suposición).

a. Calcule la estimación puntual del valor medio del espesor de recubrimiento y diga qué estimador utilizó.

b. Calcule una estimación puntual de la mediana de la distribución del espesor de recubrimiento y diga qué estimador utilizó.

c. Calcule la estimación puntual del valor que separa el 10% más grande de todos los valores de la distribución del espesor del 90% restante y diga qué estimador utilizó. [*Sugerencia*: exprese lo que está tratando de estimar en función de μ y σ.]

d. Estime $P(X < 1.5)$; es decir, la proporción de todos los valores de espesor menores que 1.5. [*Sugerencia*: si conociera los valores de μ y σ podría calcular esta probabilidad. Estos valores no están disponibles, pero pueden ser estimados.]

e. ¿Cuál es el error estándar estimado del estimador que utilizó en el inciso (b)?

4. El artículo del cual se extrajeron los datos en el ejercicio 1 también dio las observaciones de resistencias adjuntas de cilindros:

6.1 5.8 7.8 7.1 7.2 9.2 6.6 8.3 7.0 8.3
7.8 8.1 7.4 8.5 8.9 9.8 9.7 14.1 12.6 11.2

Antes de obtener los datos, denote las resistencias de vigas mediante $X_1, \ldots, X_m$ y las resistencias de cilindros $Y_1, \ldots, Y_n$. Suponga que las X_i constituyen una muestra aleatoria de una distribución con media μ_1 y desviación estándar σ_1 y que las Y_i forman una muestra aleatoria (independiente de las X_i) de otra distribución con media μ_2 y desviación estándar σ_2.

a. Use las reglas de valor esperado para demostrar que $\overline{X} - \overline{Y}$ es un estimador insesgado de $\mu_1 - \mu_2$. Calcule el estimador para los datos dados.

b. Use las reglas de varianza del capítulo 5 para obtener una expresión para la varianza y desviación estándar (error estándar) del estimador del inciso (a) y luego calcule el error estándar estimado.

c. Calcule una estimación puntual de la razón σ_1/σ_2 de las dos desviaciones estándar.

d. Suponga que se seleccionan al azar una sola viga y un solo cilindro. Calcule una estimación puntual de la varianza de la diferencia $X - Y$ entre la resistencia de las vigas y la resistencia de los cilindros.

5. Como ejemplo de una situación en la que varios estadísticos diferentes podrían ser razonablemente utilizados para calcular una estimación puntual, considere una población de N facturas. Asociado con cada factura se encuentra su "valor en libros", la cantidad anotada de dicha factura. Sea T el valor en libros total, una cantidad conocida. Algunos de estos valores en libros son erróneos. Se realizará una auditoría seleccionando al azar n facturas y determinando el valor auditado (correcto) para cada una. Suponga que la muestra da los siguientes resultados (en dólares).

	Factura				
	1	**2**	**3**	**4**	**5**
Valor en libros	300	720	526	200	127
Valor auditado	300	520	526	200	157
Error	0	200	0	0	−30

Sea

$\overline{Y}$ = valor en libros medio muestral

$\overline{X}$ = valor auditado medio muestral

$\overline{D}$ = error medio muestral

Proponga tres estadísticos diferentes para estimar el valor total (correcto) auditado: uno que implique exactamente N y $\overline{X}$, otro que implique T, N y $\overline{D}$ y el último que implique T y $\overline{X}/\overline{Y}$. Si $N = 5000$ y $T = 1{,}761{,}300$, calcule las tres estimaciones puntuales correspondientes. (El artículo "Statistical Models and Analysis in Auditing", *Statistical Science*, 1989: 2-33, discute las propiedades de estos tres estimadores.)

6. Considere las observaciones adjuntas sobre los flujos de una corriente de agua (miles de acres-pies) registradas en una estación en Colorado durante el periodo del 1 de abril al 31 de agosto durante 31 años (tomadas de un artículo que apareció en el volumen 1974 de *Water Resources Research*).

127.96 210.07 203.24 108.91 178.21
285.37 100.85 89.59 185.36 126.94
200.19 66.24 247.11 299.87 109.64
125.86 114.79 109.11 330.33 85.54
117.64 302.74 280.55 145.11 95.36
204.91 311.13 150.58 262.09 477.08
94.33

Una gráfica de probabilidad apropiada soporta el uso de la distribución lognormal (véase la sección 4.5) como modelo razonable del flujo de corriente de agua.

a. Calcule los parámetros de la distribución [*Sugerencia*: recuerde que X tiene una distribución lognormal con parámetros μ y σ^2 si $\ln(X)$ está normalmente distribuida con media μ y varianza σ^2.]

b. Use las estimaciones del inciso (a) para calcular una estimación del valor esperado del flujo de corriente de agua [*Sugerencia*: ¿Cuál es $E(X)$?]

7. a. Se selecciona una muestra aleatoria de 10 casas en un área particular, cada una de las cuales se calienta con gas natural y se determina la cantidad de gas (terms) utilizada por cada casa durante el mes de enero. Las observaciones resultantes son 103, 156, 118, 89, 125, 147, 122, 109, 138, 99. Sea μ el consumo de gas promedio durante enero de todas las casas del área. Calcule una estimación puntual de μ.

b. Suponga que hay 10,000 casas en esta área que utilizan gas natural para calefacción. Sea τ la cantidad total de gas consumido por todas estas casas durante enero. Calcule τ con los datos del inciso (a). ¿Qué estimador utilizó para calcular su estimación?

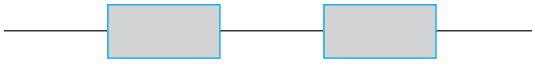
c. Use los datos del inciso (a) para estimar p, la proporción de todas las casas que usaron por lo menos 100 terms.

d. Proporcione una estimación puntual de la mediana de la población usada (el valor intermedio en la población de todas las casas) basada en la muestra del inciso (a). ¿Qué estimador utilizó?

8. En una muestra aleatoria de 80 componentes de un tipo, se encontraron 12 defectuosos.

a. Dé una estimación puntual de la proporción de todos los componentes que *no* están defectuosos.

b. Se tiene que construir un sistema seleccionando al azar dos de estos componentes y conectándolos en serie, como se muestra a continuación.



La conexión en serie implica que el sistema funcionará si y sólo si ningún componente esté defectuoso (es decir, ambos componentes funcionan apropiadamente). Estime la proporción de todos los sistemas que funcionan de manera apropiada. [*Sugerencia*: si p denota la probabilidad de que el componente funcione apropiadamente, ¿cómo puede ser expresada P(el sistema funciona) en función de p?]

9. Se examina cada uno de 150 artículos recién fabricados y se anota el número de rayones por artículo (se supone que los artículos están libres de rayones) y se obtienen los siguientes datos:

Número de rayones por artículo	0	1	2	3	4	5	6	7
Frecuencia observada	18	37	42	30	13	7	2	1

Sea X = el número de rayones en un artículo seleccionado al azar y suponga que X tiene una distribución de Poisson con parámetro μ.

a. Determine un estimador insesgado de μ y calcule la estimación de los datos. [*Sugerencia*: $E(X) = \mu$ para una distribución de Poisson de X, por lo tanto $E(\overline{X}) = ?$]

b. ¿Cuál es la desviación estándar (error estándar) de su estimador? Calcule el error estándar estimado. [*Sugerencia*: $\sigma_X^2 = \mu$ con distribución de Poisson de X.]

10. Con una larga varilla de longitud μ se va a trazar una gráfica cuadrada en la cual la longitud de cada lado es μ. Por consiguiente el área de la curva será μ^2. Sin embargo, no se conoce el valor de μ así que decide hacer n mediciones independientes $X_1, X_2, \ldots, X_n$ de la longitud. Suponga que cada X_i tiene una media μ (mediciones insesgadas) y varianza σ^2.

a. Demuestre que $\overline{X}^2$ no es un estimador insesgado de μ^2. [*Sugerencia*: con cualquier variable aleatoria Y, $E(Y^2) = V(Y) + [E(Y)]^2$. Aplique ésta con $Y = \overline{X}$.]

b. ¿Para qué valor de el estimador $\overline{X}^2 - kS^2$ k es insesgado para μ^2? [*Sugerencia*: calcule $E(\overline{X}^2 - kS^2)$.]

11. De n_1 varones fumadores seleccionados al azar, X_1 fuman cigarrillos con filtro, mientras que de n_2 fumadoras seleccionadas al azar, X_2 fuman cigarrillos con filtro. Sean p_1 y p_2 las probabilidades de que un varón y una mujer seleccionados al azar fumen, respectivamente, cigarrillos con filtro.

a. Demuestre que $(X_1/n_1) - (X_2/n_2)$ es un estimador insesgado de $p_1 - p_2$. [*Sugerencia*: $E(X_i) = n_i p_i$ con $i = 1, 2$.]

b. ¿Cuál es el error estándar del estimador en el inciso (a)?

c. ¿Cómo utilizaría los valores observados x_1 y x_2 para estimar el error estándar de su estimador?

d. Si $n_1 = n_2 = 200$, $x_1 = 127$ y $x_2 = 176$, use el estimador del inciso (a) para obtener una estimación de $p_1 - p_2$.

e. Use el resultado del inciso (c) y los datos del inciso (d) para estimar el error estándar del estimador.

12. Suponga que un tipo de fertilizante rinde μ_1 por acre con varianza σ^2, mientras que el rendimiento esperado de un segundo tipo de fertilizante es μ_2, con la misma varianza σ^2. Sean S_1^2 y S_2^2 las varianzas muestrales de los rendimientos basadas en tamaños muestrales n_1 y n_2, respectivamente, de los dos fertilizantes. Demuestre que el estimador combinado

$$\hat{\sigma}^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

es un estimador insesgado de σ^2.

13. Considere una muestra aleatoria $X_1, \ldots, X_n$ de la función de densidad de probabilidad

$$f(x; \theta) = .5(1 + \theta x) \qquad -1 \le x \le 1$$

donde $-1 \le \theta \le 1$ (esta distribución se presenta en la física de partículas). Demuestre que $\hat{\theta} = 3\overline{X}$ es un estimador insesgado de θ. [*Sugerencia*: primero determine $\mu = E(X) = E(\overline{X})$.]

14. Una muestra de n aviones de combate Pandemonium capturados tienen los números de serie $x_1, x_2, x_3, \ldots, x_n$. La CIA sabe que los aviones fueron numerados consecutivamente en la fábrica comenzando con α y terminando con β, por lo que el número total de aviones fabricados es $\beta - \alpha + 1$ (p. ej., si $\alpha = 17$ y $\beta = 29$, entonces $29 - 17 + 1 = 13$ aviones con números de serie 17, 18, 19, . . . , 28, 29 fueron fabricados). Sin embargo, la CIA no conoce los valores de α o β. Un estadístico

de la CIA sugiere utilizar el estimador máx(X_i) − mín(X_i) + 1 para estimar el número total de aviones fabricados.

a. Si $n = 5$, $x_1 = 237$, $x_2 = 375$, $x_3 = 202$, $x_4 = 525$ y $x_5 = 418$, ¿cuál es la estimación correspondiente?

b. ¿En qué condiciones de la muestra será el valor de la estimación exactamente igual al número total verdadero de aviones? ¿Alguna vez será más grande la estimación que el total verdadero? ¿Piensa que el estimador es insesgado para estimar $\beta - \alpha + 1$? Explique en una o dos oraciones.

15. Si $X_1, X_2, \ldots, X_n$ representan una muestra aleatoria tomada de una distribución de Rayleigh con función de densidad de probabilidad

$$f(x;\theta) = \frac{x}{\theta} e^{-x^2/(2\theta)} \quad x > 0$$

a. Se puede demostrar que $E(X^2) = 2\theta$. Use este hecho para construir un estimador insesgado de θ basado en $\sum X_i^2$ (y use reglas de valor esperado para demostrar que es insesgado).

b. Calcule θ a partir de las siguientes $n = 10$ observaciones de esfuerzo vibratorio de un aspa de turbina en condiciones específicas:

16.88	10.23	4.59	6.66	13.68
14.23	19.87	9.40	6.51	10.95

16. Suponga que el crecimiento promedio verdadero μ de un tipo de planta durante un periodo de 1 año es idéntico al de un segundo tipo, aunque la varianza del crecimiento del primer tipo es σ^2, en tanto que para el segundo tipo la varianza es $4\sigma^2$. Sean $X_1, \ldots, X_m$, m observaciones de crecimiento independientes del primer tipo [por consiguiente $E(X_i) = \mu$, $V(X_i) = \sigma^2$], y sean $Y_1, \ldots, Y_n$, n observaciones de crecimiento independientes del segundo tipo [$E(Y_i) = \mu$, $V(Y_i) = 4\sigma^2$].

a. Demuestre que para cualquier δ entre 0 y 1, el estimador $\hat{\mu} = \delta\overline{X} + (1 - \delta)\overline{Y}$ es insesgado para μ.

b. Con m y n fijas, calcule $V(\hat{\mu})$ y luego determine el valor de δ que reduzca al mínimo $V(\hat{\mu})$. [*Sugerencia*: diferencie $V(\hat{\mu})$ con respecto a δ.]

17. En el capítulo 3 se definió una variable aleatoria binomial negativa como el número de fallas que ocurren antes del r-ésimo éxito en una secuencia de ensayos con éxitos y fallos independientes e idénticos. La función de masa de probabilidad (fmp) de X es

$$nb(x; r, p) = \binom{x + r - 1}{x} p^r(1 - p)^x \quad x = 0, 1, 2, \ldots$$

a. Suponga que $r \geq 2$. Demuestre que

$$\hat{p} = (r - 1)/(X + r - 1)$$

es un estimador insesgado de p. [*Sugerencia*: escriba $E(\hat{p})$ y elimine $x + r - 1$ dentro de la suma.]

b. Un reportero desea entrevistar a cinco individuos que apoyan a un candidato y comienza preguntándoles si (S) o no (F) apoyan al candidato. Si la secuencia de respuestas es *SFFSFFFSSS*, estime p = la proporción verdadera que apoya al candidato.

18. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una función de densidad de probabilidad $f(x)$ que es simétrica con respecto a μ, de modo que $\widetilde{X}$ es un estimador insesgado de μ. Si n es grande, se puede demostrar que $V(\widetilde{X}) \approx 1/(4n[f(\mu)]^2)$.

a. Compara $V(\widetilde{X})$ con $V(\overline{X})$ cuando la distribución subyacente es normal.

b. Cuando la función de densidad de probabilidad subyacente es de Cauchy (véase el ejemplo 6.7), $V(\overline{X}) = \infty$, por lo tanto $\overline{X}$ es un estimador terrible. ¿Cuál es $V(\widetilde{X})$ en este caso cuando n es grande?

19. Una investigadora desea estimar la proporción de estudiantes en una universidad que han violado el código de honor. Habiendo obtenido una muestra aleatoria de n estudiantes, se da cuenta de que si a cada uno le pregunta "¿has violado el código de honor?" probablemente recibirá algunas respuestas faltas de veracidad. Considere el siguiente esquema, conocido como técnica de **respuesta aleatorizada.** La investigadora forma un mazo de 100 cartas de las cuales 50 son de tipo I y 50 de tipo II.

Tipo I: ¿Has violado el código de honor (sí o no)?

Tipo II: ¿El último dígito de su número telefónico es un 0, 1 o 2 (sí o no)?

A cada estudiante en la muestra aleatoria se le pide que baraje el mazo, que saque una carta y que responda la pregunta con sinceridad. A causa de la pregunta irrelevante en las cartas de tipo II, una respuesta afirmativa ya no estigmatiza al que responde, así que se supone que éste es sincero. Sea p la proporción de violadores del código de honor (es decir, la probabilidad de que un estudiante seleccionado al azar sea un violador) y sea $\lambda = P$(respuesta sí). Entonces λ y p están relacionados por $\lambda = .5p + (.5)(.3)$.

a. Sea Y el número de respuestas afirmativa, por consiguiente $Y \sim$ Bin (n, λ). Por tanto Y/n es un estimador insesgado de λ. Deduzca un estimador de p basado en Y. Si $n = 80$ y $y = 20$, ¿cuál es su estimación? [*Sugerencia*: despeje p de $\lambda = .5p + .15$ y luego sustituya Y/n en lugar de λ.]

b. Use el hecho de que $E(Y/n) = \lambda$ para demostrar que su estimador $\hat{p}$ es insesgado.

c. Si hubiera 70 cartas de tipo I y 30 de tipo II, ¿cuál sería su estimador para p?

6.2 Métodos de estimación puntual

La definición de ausencia de sesgo no indica en general cómo se pueden obtener los estimadores insesgados. A continuación se discuten dos métodos "constructivos" para obtener estimadores puntuales: el método de momentos y el método de máxima probabilidad. Por *constructivo* se quiere dar a entender que la definición general de cada tipo de estimador

sugiere explícitamente cómo obtener el estimador en cualquier problema específico. Aun cuando se prefieren los estimadores de máxima probabilidad a los de momento debido a ciertas propiedades de eficiencia, a menudo requieren significativamente más cálculo que los estimadores de momento. En ocasiones es el caso que estos métodos dan estimadores insesgados.

El método de momentos

La idea básica de este método es poder igualar ciertas características muestrales, tales como la media, a los valores esperados de la población correspondiente. Luego al resolver estas ecuaciones para valores de parámetros desconocidos se obtienen los estimadores.

DEFINICIÓN

Si $X_1, \ldots, X_n$ constituyen una muestra aleatoria proveniente de una función de masa de probabilidad o de una función de densidad de probabilidad $f(x)$. Con $k = 1, 2, 3, \ldots$, el **momento *k*-ésimo de la población** o el **momento *k*-ésimo de la distribución *f*(*x*),** es $E(X^k)$. El **momento muestral *k*-ésimo** es $(1/n)\sum_{i=1}^{n} X_i^k$.

Por consiguiente, el primer momento de la población es $E(X) = \mu$ y el primer momento muestral es $\Sigma X_i/n = \overline{X}$. Los segundos momentos de la población y muestral son $E(X^2)$ y $\sum X_i^2/n$, respectivamente. Los momentos de la población serán funciones de cualesquiera parámetros desconocidos $\theta_1, \theta_2, \ldots$.

DEFINICIÓN

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución con función de masa de probabilidad o función de densidad de probabilidad $f(x; \theta_1, \ldots, \theta_m)$, donde $\theta_1, \ldots, \theta_m$ son parámetros cuyos valores son desconocidos. Entonces los **estimadores de momento** $\hat{\theta}_1, \ldots, \hat{\theta}_m$ se obtienen igualando los primeros m momentos muestrales con los primeros m momentos de la población correspondientes y resolviendo para $\theta_1, \ldots, \theta_m$.

Si, por ejemplo, $m = 2$, $E(X)$ y $E(X^2)$ serán funciones de θ_1 y θ_2. Con $E(X) = (1/n)\,\Sigma X_i$ $(= \overline{X})$ y $E(X^2) = (1/n)\sum X_i^2$ se obtienen dos ecuaciones con θ_1 y θ_2. La solución define entonces los estimadores.

Ejemplo 6.12 Si $X_1, X_2, \ldots, X_n$ representan una muestra aleatoria de tiempos de servicio de n clientes en una instalación, donde la distribución subyacente se supone exponencial con el parámetro λ. Como sólo hay un parámetro que tiene que ser estimado, el estimador se obtiene igualando $E(X)$ a $\overline{X}$. Como $E(X) = 1/\lambda$ con una distribución exponencial, ésta da $1/\lambda = \overline{X}$ o $\lambda = 1/\overline{X}$. El estimador de momento de λ es entonces $\hat{\lambda} = 1/\overline{X}$. ■

Ejemplo 6.13 Sean $X_1, \ldots, X_n$ una muestra aleatoria de una distribución gamma con parámetros α y β. De acuerdo con la sección 4.4, $E(X) = \alpha\beta$ y $E(X^2) = \beta^2\Gamma(\alpha + 2)/\Gamma(\alpha) = \beta^2(\alpha + 1)\alpha$. Los estimadores de momento de α y β se obtienen resolviendo

$$\overline{X} = \alpha\beta \qquad \frac{1}{n}\sum X_i^2 = \alpha(\alpha + 1)\beta^2$$

Como $\alpha(\alpha + 1)\beta^2 = \alpha^2\beta^2 + \alpha\beta^2$ y la primera ecuación implica $\alpha^2\beta^2 = \overline{X}^2$, la segunda ecuación se vuelve

$$\frac{1}{n}\sum X_i^2 = \overline{X}^2 + \alpha\beta^2$$

Ahora si se divide cada miembro de esta segunda ecuación entre el miembro correspondiente de la primera ecuación y se sustituye otra vez se obtienen los estimadores

$$\hat{\alpha} = \frac{\overline{X}^2}{(1/n)\sum X_i^2 - \overline{X}^2} \qquad \hat{\beta} = \frac{(1/n)\sum X_i^2 - \overline{X}^2}{\overline{X}}$$

Para ilustrar, los datos de tiempo de sobrevivencia mencionados en el ejemplo 4.24 son

152	115	109	94	88	137	152	77	160	165
125	40	128	123	136	101	62	153	83	69

con $\bar{x} = 113.5$ y $(1/20)\sum x_i^2 = 14{,}087.8$. Los estimadores son

$$\hat{\alpha} = \frac{(113.5)^2}{14{,}087.8 - (113.5)^2} = 10.7 \qquad \hat{\beta} = \frac{14{,}087.8 - (113.5)^2}{113.5} = 10.6$$

Estas estimaciones de α y β difieren de los valores sugeridos por Gross y Clark porque ellos utilizaron una técnica de estimación diferente. ■

Ejemplo 6.14 Sean $X_1, \ldots, X_n$ una muestra aleatoria de una distribución binomial negativa generalizada con parámetros r y p (vea la sección 3.5). Como $E(X) = r(1-p)/p$ y $V(X) = r(1-p)/p^2$, $E(X^2) = V(X) + [E(X)]^2 = r(1-p)(r - rp + 1)/p^2$. Si se iguala $E(X)$ a $\overline{X}$ y $E(X^2)$ a $(1/n)\sum X_i^2$ a la larga se obtiene

$$\hat{p} = \frac{\overline{X}}{(1/n)\sum X_i^2 - \overline{X}^2} \qquad \hat{r} = \frac{\overline{X}^2}{(1/n)\sum X_i^2 - \overline{X}^2 - \overline{X}}$$

Como ilustración, Reep, Pollard y Benjamin (“Skill and Chance in Ball Games”, *J. of Royal Stat. Soc.*, 1971: 623–629) consideran la distribución binomial negativa como modelo del número de goles por juego anotados por los equipos de la Liga Nacional de Jockey. Los datos de 1966–1967 son los siguientes (420 juegos):

Goles	0	1	2	3	4	5	6	7	8	9	10
Frecuencia	29	71	82	89	65	45	24	7	4	1	3

Entonces,

$$\bar{x} = \Sigma x_i/420 = [(0)(29) + (1)(71) + \cdots + (10)(3)]/420 = 2.98$$

y

$$\Sigma x_i^2/420 = [(0)^2(29) + (1)^2(71) + \cdots + (10)^2(3)]/420 = 12.40$$

Por consiguiente,

$$\hat{p} = \frac{2.98}{12.40 - (2.98)^2} = .85 \qquad \hat{r} = \frac{(2.98)^2}{12.40 - (2.98)^2 - 2.98} = 16.5$$

Aunque r por definición debe ser positivo, el denominador de $\hat{r}$ podría ser negativo, lo que indica que la distribución binomial negativa no es apropiada (o que el estimador de momento es defectuoso). ■

Estimación de máxima probabilidad

El método de máxima probabilidad lo introdujo por primera vez R. A. Fisher, genetista y estadístico en la década de 1920. La mayoría de los estadísticos recomiendan este método, por lo menos cuando el tamaño de muestra es grande, puesto que los estimadores resultantes tienen ciertas propiedades de eficiencia deseables (véase la proposición en la página 262).

Ejemplo 6.15 Se obtuvo una muestra de diez cascos de ciclista nuevos fabricados por una compañía. Al probarlos, se encontró que el primero, el tercero y el décimo estaban agrietados, en tanto que los demás no. Sea $p = P(\text{casco agrietado})$ es decir, p es la proporción de todos los cascos que están agrietados. Defina variables aleatorias (de Bernoulli) $X_1, X_2, \ldots, X_{10}$ como

$$X_1 = \begin{cases} 1 \text{ si el primer casco está agrietado} \\ 0 \text{ si el primer casco no está agrietado} \end{cases} \cdots X_{10} = \begin{cases} 1 \text{ si el décimo casco está agrietado} \\ 0 \text{ si el décimo casco no está agrietado} \end{cases}$$

Entonces, para la muestra obtenida, $X_1 = X_3 = X_{10} = 1$ y las otras siete X_i son cero. La función de masa de probabilidad de cualquier X_i particular es $p^{x_i}(1-p)^{1-x_i}$ que se convierte en p si $x_i = 1$ y $1 - p$ cuando $x_i = 0$. Supóngase ahora que las condiciones de los diferentes cascos son independientes una de la otra. Esto implica que las X_i son independientes, por lo que su función de masa de probabilidad conjunta es el producto de las funciones de masa de probabilidad individuales. Así que la función de masa de probabilidad conjunta de las X_i observadas es

$$f(x_1, \ldots, x_{10}; p) = p(1-p)p \cdots p = p^3(1-p)^7 \tag{6.4}$$

Supóngase que $p = .25$. Entonces la probabilidad de observar la muestra que en realidad se obtiene es $(.25)^3\,(.75)^7 = .002086$. Si en cambio $p = .50$, entonces esta probabilidad es $(.50)^3\,(.50)^7 = .000977$. ¿Para qué valor de p es más probable que la muestra observada haya ocurrido? Es decir, ¿para qué valor de p es la función de masa de probabilidad conjunta (6.4) tan grande como puede ser? ¿Qué valor de p maximiza (6.4)? La figura 6.5(a) muestra un gráfico de la *probabilidad* (6.4) en función de p. Parece que la gráfica alcanza su pico por encima de $p = .3 =$ la proporción de los cascos defectuosos en la muestra. La figura 6.5(b) muestra un gráfico del logaritmo natural de (6.4) ya que $ln[g(u)]$ es una función estrictamente creciente de $g(u)$, encontrar u para maximizar la función $g(u)$ es lo mismo que encontrar u para maximizar $ln[g(u)]$.

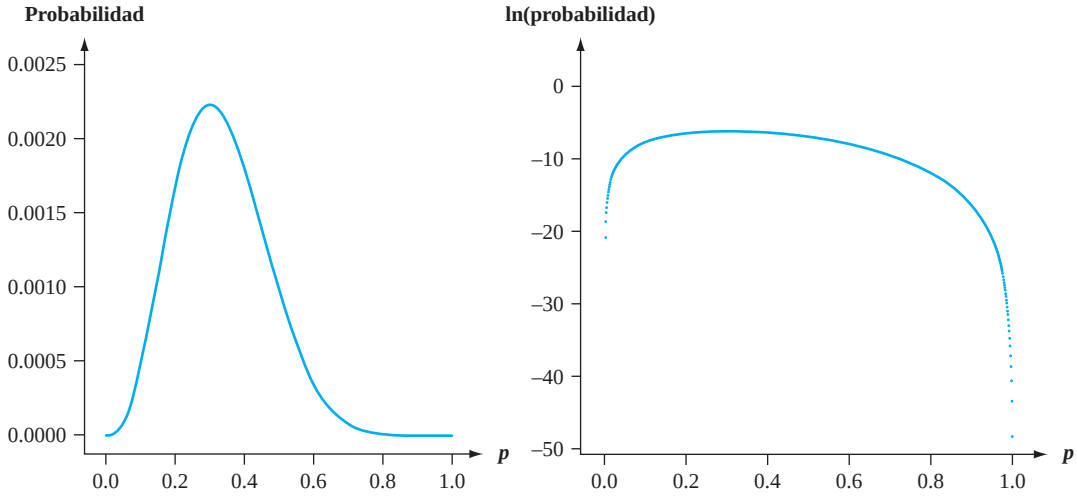


Figura 6.5 (a) Gráfica de la probabilidad (función de masa de probabilidad conjunta) (6.4) del ejemplo 6.15; (b) Gráfica del logaritmo natural de la probabilidad

Podemos comprobar nuestra impresión visual usando el cálculo para hallar el valor de p que maximiza (6.4). Trabajar con el logaritmo natural de la función de masa de probabilidad conjunta suele ser más fácil que trabajar con la función de masa de probabilidad conjunta, como esta última es típicamente un producto, su logaritmo será una suma. Aquí

$$\ln[\, f(x_1, \ldots, x_{10}; p)] = \ln[p^3(1-p)^7] = 3\ln(p) + 7\ln(1-p) \tag{6.5}$$

Por tanto

$$\frac{d}{dp}\{\ln[f(x_1, \ldots, x_{10}; p)]\} = \frac{d}{dp}\{3\ln(p) + 7\ln(1-p)\} = \frac{3}{p} + \frac{7}{1-p}(-1)$$
$$= \frac{3}{p} - \frac{7}{1-p}$$

[el (-1) viene de la regla de la cadena en el cálculo]. Igualando esta derivada a 0 y despejando p da $3(1-p) = 7p$, de lo cual $3 = 10p$ y así $p = 3/10 = .30$ como conjeturamos. Es decir, nuesta estimación puntual es $\hat{p} = .30$. Se llama *estimación de máxima verosimilitud*, ya que es el valor del parámetro que maximiza la probabilidad (fmp conjunta) de la muestra observada. En general, la segunda derivada debe ser examinada para asegurarse de que se haya obtenido un máximo, pero aquí esto es obvio en la figura 6.5.

Supongamos que en vez de decirle la condición de cada casco, sólo había sido informado de que tres de los diez eran defectuosos. Entonces tendríamos el valor observado de una variable aleatoria binomial $X =$ el número de cascos defectuosos. La función de masa de probabilidad de X es $\binom{10}{x} p^x (1-p)^{10-x}$. Para $x = 3$, esto se convierte en $\binom{10}{3} p^3 (1-p)^7$. El coeficiente binomial $\binom{10}{3}$ es irrelevante para la maximización, así que nuevamente $\hat{p} = .30$. ■

DEFINICIÓN

Sean $X_1, X_2, \ldots, X_n$ que tienen una función de masa de probabilidad o una función de densidad de probabilidad

$$f(x_1, x_2, \ldots, x_n; \theta_1, \ldots, \theta_m) \tag{6.6}$$

donde los parámetros $\theta_1, \ldots, \theta_m$ tienen valores desconocidos. Cuando $x_1, \ldots, x_n$ son los valores muestrales observados y (6.6) se considera como una función de $\theta_1, \ldots, \theta_m$, se llama **función de probabilidad.** Las estimaciones de máxima probabilidad $\hat{\theta}_1, \ldots, \hat{\theta}_m$ son aquellos valores de las θ_i que incrementan al máximo la función de probabilidad, de modo que

$$f(x_1, \ldots, x_n; \hat{\theta}_1, \ldots, \hat{\theta}_m) \geq f(x_1, \ldots, x_n; \theta_1, \ldots, \theta_m) \text{ para todas las } \theta_1, \ldots, \theta_m$$

Cuando se sustituyen las X_i en lugar de las x_i, se obtienen los **estimadores de máxima probabilidad**.

La función de probabilidad dice qué tan probable es que la muestra observada sea una función de los posibles valores de parámetro. Al incrementarse al máximo la probabilidad se obtienen los valores de parámetro con los que es más probable que la muestra observada haya sido generada; es decir, los valores de parámetro que "más concuerdan" con los datos observados.

Ejemplo 6.16 Suponga que $X_1, X_2, \ldots, X_n$ es una muestra aleatoria de una distribución exponencial con parámetro λ. Debido a la independencia, la función de probabilidad es un producto de las funciones de densidad de probabilidad individuales:

$$f(x_1, \ldots, x_n; \lambda) = (\lambda e^{-\lambda x_1}) \cdot \cdots \cdot (\lambda e^{-\lambda x_n}) = \lambda^n e^{-\lambda \Sigma x_i}$$

El logaritmo natural de la función de probabilidad es

$$\ln[f(x_1, \ldots, x_n; \lambda)] = n\ln(\lambda) - \lambda \Sigma x_i$$

Si se iguala $(d/d\lambda)$[ln(probabilidad)] a cero se obtiene $n/\lambda - \Sigma x_i = 0$ o $\lambda = n/\Sigma x_i = 1/\bar{x}$. Por consiguiente el estimador de máxima probabilidad es $\hat{\lambda} = 1/\bar{X}$; es idéntico al método de estimador de momentos [pero no es un estimador insesgado, puesto que $E(1/\bar{X}) \neq 1/E(\bar{X})$]. ■

Ejemplo 6.17 Sean $X_1, \ldots, X_n$ una muestra aleatoria de una distribución normal. La función de probabilidad es

$$f(x_1, \ldots, x_n; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_1-\mu)^2/(2\sigma^2)} \cdot \cdots \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_n-\mu)^2/(2\sigma^2)}$$
$$= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} e^{-\Sigma(x_i-\mu)^2/(2\sigma^2)}$$

por consiguiente

$$\ln[f(x_1, \ldots, x_n; \mu, \sigma^2)] = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \Sigma(x_i - \mu)^2$$

Para determinar los valores maximizantes de μ y σ^2, se deben sacar las derivadas parciales de $\ln(f)$ con respecto a μ y σ^2, igualarlas a cero y resolver las dos ecuaciones resultantes. Omitiendo los detalles, los estimadores de máxima probabilidad resultantes son

$$\hat{\mu} = \overline{X} \qquad \hat{\sigma}^2 = \frac{\Sigma(X_i - \overline{X})^2}{n}$$

El estimador de máxima probabilidad de σ^2 no es el estimador insesgado, por consiguiente dos principios diferentes de estimación (ausencia de sesgo y máxima probabilidad) dan dos estimadores diferentes. ■

Ejemplo 6.18 En el capítulo 3 se mencionó el uso de la distribución de Poisson para modelar el número de "eventos" que ocurren en una región bidimensional. Suponga que cuando la región R se está muestreando tiene área $a(R)$, el número X de eventos que ocurren en R tiene una distribución de Poisson con parámetro $\lambda a(R)$ (donde λ es el número esperado de eventos por unidad de área) y que las regiones no traslapantes dan X independientes.

Suponga que un ecólogo selecciona n regiones no traslapantes $R_1, \ldots, R_n$ y cuenta el número de plantas de una especie en cada región. La función de masa de probabilidad (verosimilitud) conjunta es entonces

$$p(x_1, \ldots, x_n; \lambda) = \frac{[\lambda \cdot a(R_1)]^{x_1} e^{-\lambda \cdot a(R_1)}}{x_1!} \cdot \cdots \cdot \frac{[\lambda \cdot a(R_n)]^{x_n} e^{-\lambda \cdot a(R_n)}}{x_n!}$$
$$= \frac{[a(R_1)]^{x_1} \cdot \cdots \cdot [a(R_n)]^{x_n} \cdot \lambda^{\Sigma x_i} \cdot e^{-\lambda \Sigma a(R_i)}}{x_1! \cdot \cdots \cdot x_n!}$$

El ln(probabilidad) es

$$\ln[p(x_1, \ldots, x_n; \lambda)] = \Sigma x_i \cdot \ln[a(R_i)] + \ln(\lambda) \cdot \Sigma x_i - \lambda \Sigma a(R_i) - \Sigma \ln(x_i!)$$

Tomando $d/d\lambda \ln(p)$ e igualándola a cero da

$$\frac{\Sigma x_i}{\lambda} - \Sigma a(R_i) = 0$$

por consiguiente

$$\lambda = \frac{\Sigma x_i}{\Sigma a(R_i)}$$

El estimador de máxima probabilidad es entonces $\hat{\lambda} = \Sigma X_i / \Sigma a(R_i)$. Ésta es razonablemente intituitiva porque λ es la densidad verdadera (plantas por unidad de área), mientras que $\hat{\lambda}$ es la densidad muestral puesto que $\Sigma a(R_i)$ es tan sólo el área total muestreada. Como $E(X_i) = \lambda \cdot a(R_i)$, el estimador es insesgado.

En ocasiones se utiliza un procedimiento de muestreo alternativo. En lugar de fijar las regiones que van a ser muestreadas, el ecólogo seleccionará n puntos en toda la región

de interés y sea y_i = la distancia del punto i-ésimo a la planta más cercana. La función de distribución acumulativa de Y = distancia a la planta más cercana es

$$F_Y(y) = P(Y \le y) = 1 - P(Y > y) = 1 - P\begin{pmatrix}\text{ninguna planta en}\\ \text{un círculo de radio } y\end{pmatrix}$$

$$= 1 - \frac{e^{-\lambda\pi y^2}(\lambda\pi y^2)^0}{0!} = 1 - e^{-\lambda \cdot \pi y^2}$$

Al sacar la derivada de $F_Y(y)$ con respecto a y resulta

$$f_Y(y;\lambda) = \begin{cases} 2\pi\lambda y e^{-\lambda\pi y^2} & y \ge 0 \\ 0 & \text{de lo contrario} \end{cases}$$

Si ahora se forma la probabilidad $f_Y(y_1;\lambda) \cdot \cdots \cdot f_Y(y_n;\lambda)$, se diferencia ln(probabilidad), y así sucesivamente, el estimador de máxima probabilidad resultante es

$$\hat{\lambda} = \frac{n}{\pi\sum Y_i^2} = \frac{\text{número de plantas observadas}}{\text{área total muestreada}}$$

la que también es una densidad muestral. Se puede demostrar que un ambiente ralo (pequeño λ), el método de distancia es mejor en cierto sentido, en tanto que en un ambiente denso, el primer método de muestreo es mejor. ■

Ejemplo 6.19 Sean $X_1, \ldots, X_n$ una muestra aleatoria de una función de densidad de probabilidad de Weibull

$$f(x;\alpha,\beta) = \begin{cases} \dfrac{\alpha}{\beta^\alpha} \cdot x^{\alpha-1} \cdot e^{-(x/\beta)^\alpha} & x \ge 0 \\ 0 & \text{de lo contrario} \end{cases}$$

Si se escribe la probabilidad y el ln(probabilidad) y luego $(\partial/\partial\alpha)[\ln(f)] = 0$ y $(\partial/\partial\beta)[\ln(f)] = 0$ se obtienen las ecuaciones

$$\alpha = \left[\frac{\sum x_i^\alpha \cdot \ln(x_i)}{\sum x_i^\alpha} - \frac{\sum \ln(x_i)}{n}\right]^{-1} \qquad \beta = \left(\frac{\sum x_i^\alpha}{n}\right)^{1/\alpha}$$

Estas dos ecuaciones no pueden ser resueltas explícitamente para obtener fórmulas generales de los estimadores de máxima probabilidad $\hat{\alpha}$ y $\hat{\beta}$. En su lugar, por cada muestra $x_1, \ldots, x_n$, las ecuaciones deben ser resueltas con un procedimiento numérico iterativo. Incluso los estimadores de momento de α y β son un tanto complicados (véase el ejercicio 21). ■

Estimación de funciones de parámetros

En el ejemplo 6.17, se obtuvo el estimador de máxima probabilidad de σ^2 cuando la distribución subyacente es normal. El estimador de máxima probabilidad de $\sigma = \sqrt{\sigma^2}$ como el de muchos otros estimadores de máxima probabilidad, es fácil de obtener con la siguiente proposición.

PROPOSICIÓN

El principio de invarianza

Sean $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m$ los estimadores de máxima probabilidad de los parámetros $\theta_1, \theta_2, \ldots, \theta_m$. Entonces el estimador de máxima probabilidad de cualquier función $h(\theta_1, \theta_2, \ldots, \theta_m)$ de estos parámetros es la función $h(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m)$ de los estimadores de máxima probabilidad.

Ejemplo 6.20 (Continuación del ejemplo 6.17)

En el caso normal, los estimadores de máxima probabilidad de μ y σ^2 son $\hat{\mu} = \overline{X}$ y $\hat{\sigma}^2 = \sum(X_i - \overline{X})^2/n$. Para obtener el estimador de máxima probabilidad de la función $h(\mu, \sigma^2) = \sqrt{\sigma^2} = \sigma$, sustituya los estimadores de máxima probabilidad en la función:

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \left[\frac{1}{n}\sum(X_i - \overline{X})^2\right]^{1/2}$$

El estimador de máxima probabilidad de σ no es la desviación estándar muestral S, aunque se aproximan bastante a menos que n sea bastante pequeño. ■

Ejemplo 6.21 (Continuación del ejemplo 6.19)

El valor medio de una variable aleatoria X que tiene una distribución de Weibull es

$$\mu = \beta \cdot \Gamma(1 + 1/\alpha)$$

El estimador de máxima probabilidad de μ es por consiguiente $\hat{\mu} = \hat{\beta}\Gamma(1 + 1/\hat{\alpha})$ donde $\hat{\alpha}$ y $\hat{\beta}$ son los estimadores de máxima probabilidad de α y β. En particular, $\overline{X}$ no es el estimador de máxima probabilidad de μ, aunque es un estimador insesgado. Por lo menos para n grande, $\hat{\mu}$ es un mejor estimador que $\overline{X}$.

Para los datos que figuran en el ejemplo 6.3, los estimadores de máxima probabilidad de los parámetros de Weibull son $\hat{\alpha} = 11.9731$ y $\hat{\beta} = 77.0153$, de los cuales $\hat{\mu} = 73.80$. Esta estimación está muy cerca de la media de la muestra 73.88. ■

Comportamiento del estimador de máxima probabilidad con muestra grande

Aunque el principio de la estimación de máxima probabilidad tiene un considerable atractivo intuitivo, la siguiente proposición proporciona razones adicionales fundamentales para el uso de estimadores de máxima probabilidad.

PROPOSICIÓN

En condiciones muy generales en relación con la distribución conjunta de la muestra, cuando el tamaño n de la muestra es grande, el estimador de máxima probabilidad de cualquier parámetro θ es aproximadamente insesgado $[E(\hat{\theta}) \approx \theta]$ y su varianza es casi tan pequeña como la que puede ser lograda por cualquier estimador. Expresado de otra manera, el estimador de máxima probabilidad $\hat{\theta}$ es aproximadamente el estimador insesgado con varianza mínima de θ.

Debido a este resultado y al hecho de que las técnicas basadas en el cálculo casi siempre pueden ser utilizadas para obtener los estimadores de máxima probabilidad (aunque a veces se requieren métodos numéricos, tales como el método de Newton), la estimación de máxima probabilidad es la técnica de estimación más ampliamente utilizada entre los estadísticos. Muchos de los estimadores utilizados en lo que resta del libro son estimadores de máxima probabilidad. Sin embargo, la obtención de un estimador de máxima probabilidad requiere que se especifique la distribución subyacente.

Algunas complicaciones

En ocasiones no se puede utilizar el cálculo para obtener estimadores de máxima probabilidad.

Ejemplo 6.22

Suponga que mi tiempo de espera de un autobús está uniformemente distribuido en $[0, \theta]$ y que se observaron los resultados $x_1, \ldots, x_n$ de una muestra aleatoria tomada de esta distribución. Como $f(x; \theta) = 1/\theta$ con $0 \le x \le \theta$ y 0 de lo contrario,

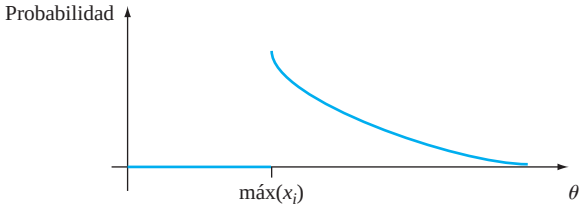


Figura 6.6 Función de probabilidad para el ejemplo 6.22

$$f(x_1, \ldots, x_n; \theta) = \begin{cases} \dfrac{1}{\theta^n} & 0 \le x_1 \le \theta, \ldots, 0 \le x_n \le \theta \\ 0 & \text{de lo contrario} \end{cases}$$

En tanto máx(x_i) $\le \theta$, la probabilidad es $1/\theta^n$, la cual es positiva, pero en cuanto $\theta <$ máx(x_i), la probabilidad se reduce a 0. Esto se ilustra en la figura 6.6. El cálculo no funciona porque el máximo de la probabilidad ocurre en un punto de discontinuidad, pero la figura indica que $\hat{\theta} =$ máx(X_i). Por consiguiente si mis tiempos de espera son 2.3, 3.7, 1.5, .4 y 3.2, entonces el estimador de máxima probabilidad es $\hat{\theta} = 3.7$. De acuerdo con el ejemplo 6.4, el estimador de máxima probabilidad no es insesgado. ■

Ejemplo 6.23 Un método que a menudo se utiliza para estimar el tamaño de una población de vida silvestre implica realizar un experimento de captura-recaptura. En este experimento se captura una muestra inicial de M animales, cada uno se éstos se etiqueta y luego son regresados a la población. Tras permitir un tiempo suficiente para que los individuos etiquetados se mezclen con la población, se captura otra muestra de tamaño n. Con $X =$ el número de animales etiquetados en la segunda muestra, el objetivo es utilizar las x observadas para estimar la población de tamaño N.

El parámetro de interés es $\theta = N$, el cual puede asumir sólo valores enteros, así que incluso después de determinar la función de probabilidad (función de masa de probabilidad de X en este caso), el uso del cálculo para obtener N presentaría dificultades. Si se considera un éxito la recaptura de un animal previamente etiquetado, entonces el muestreo es sin reemplazo de una población que contiene M éxitos y $N - M$ fallas, de modo que X es una variable aleatoria hipergeométrica y la función de probabilidad es

$$p(x; N) = h(x; n, M, N) = \frac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}}$$

No obstante la naturaleza de valor entero de N, sería difícil evaluar la derivada de $p(x; N)$. Sin embargo, si se considera la razón de $p(x; N)$ y $p(x; N - 1)$, se tiene

$$\frac{p(x; N)}{p(x; N-1)} = \frac{(N-M) \cdot (N-n)}{N(N-M-n+x)}$$

Esta razón es más grande que 1 si y sólo si $N < Mn/x$. El valor de N con el cual $p(x; N)$ se incrementa al máximo es por consiguiente el entero más grande menor que Mn/x. Si se utiliza la notación matemática estándar $[r]$ para el entero más grande menor o igual a r, el estimador de máxima probabilidad de N es $\hat{N} = [Mn/x]$. Como ilustración, si $M = 200$ peces se sacan de un lago y etiquetan, posteriormente $n = 100$ son recapturados y entre los 100 hay $x = 11$ etiquetados, en ese caso $\hat{N} = [(200)(100)/11] = [1818.18] = 1818$. La estimación es en realidad un tanto intuitiva; x/n es la proporción de la muestra recapturada etiquetada, mientras que M/N es la proporción de toda la población etiquetada. La estimación se obtiene igualando estas dos proporciones (estimando una proporción poblacional mediante una proporción muestral). ■

Supóngase que $X_1, X_2, \ldots, X_n$ es una muestra aleatoria de una función de densidad de probabilidad $f(x; \theta)$ simétrica con respecto a θ aunque el investigador no está seguro de la forma de la función f. Es entonces deseable utilizar un estimador $\hat{\theta}$ *robusto*; es decir, uno que funcione bien con una amplia variedad de funciones de densidad de probabilidad subyacentes. Un estimador como ése es una media recortada. En años recientes, los estadísticos han propuesto otro tipo de estimador, llamado *estimador M*, basado en una generalización de la estimación de máxima probabilidad. En lugar de incrementar al máximo el logaritmo de la probabilidad $\Sigma \ln[f(x; \theta)]$ para una f específica, se incrementa al máximo $\Sigma \rho(x_i; \theta)$. Se selecciona la "función objetivo" ρ para que dé un estimador con buenas propiedades de robustez. El libro de David Hoaglin y colaboradores (véase la bibliografía) contiene una buena exposición de esta materia.

EJERCICIOS Sección 6.2 (20–30)

20. Una prueba de diagnóstico para una enfermedad determinada se aplica a n individuos de los que se sabe que no tienen la enfermedad. Sea X = el número uno de los n resultados de prueba que son positivos (lo que indica la presencia de la enfermedad, por lo que X es el número de falsos positivos) y p = probabilidad de que el resultado de un individuo de prueba libre de la enfermedad es positivo (es decir, p es la verdadera proporción de resultados de las pruebas de individuos libres de enfermedades que son positivos). Supongamos que sólo X está disponible en lugar de la secuencia real de los resultados de la prueba.

a. Derive el estimador de máxima probabilidad de p. Si $n = 20$ y $x = 3$, ¿cuál es la estimación?

b. ¿Es insesgado el estimador del inciso (a)?

c. Si $n = 20$ y $x = 3$, ¿cuál es el estimador de máxima probabilidad $(1 - p)^5$ que ninguna de las próximas cinco pruebas realizadas en los individuos libres de la enfermedad sean positivas?

21. Si X tiene una distribución de Weibull con parámetros α y β, entonces

$$E(X) = \beta \cdot \Gamma(1 + 1/\alpha)$$
$$V(X) = \beta^2\{\Gamma(1 + 2/\alpha) - [\Gamma(1 + 1/\alpha)]^2\}$$

a. Basado en una muestra aleatoria $X_1, \ldots, X_n$, escriba ecuaciones para el método de estimadores de momentos de β y α. Demuestre que una vez que se obtiene la estimación de α, la estimación de β se puede hallar en una tabla de la función gamma, y que la estimación de α es la solución de una ecuación complicada que implica la función gamma.

b. Si $n = 20$, $\bar{x} = 28.0$ y $\sum x_i^2 = 16{,}500$, calcule la estimaciones. [*Sugerencia*: $[\Gamma(1.2)]^2/\Gamma(1.4) = .95$.]

22. Sea X la proporción de tiempo asignado que un estudiante seleccionado al azar pasa resolviendo cierta prueba de aptitud. Suponga que la función de densidad de probabilidad de X es

$$f(x; \theta) = \begin{cases} (\theta + 1)x^\theta & 0 \le x \le 1 \\ 0 & \text{de lo contrario} \end{cases}$$

donde $-1 < \theta$. Una muestra aleatoria de diez estudiantes produce los datos $x_1 = .92$, $x_2 = .79$, $x_3 = .90$, $x_4 = .65$, $x_5 = .86$, $x_6 = .47$, $x_7 = .73$, $x_8 = .97$, $x_9 = .94$, $x_{10} = .77$.

a. Use el método de momentos para obtener un estimador de θ y luego calcule la estimación para estos datos.

b. Obtenga el estimador de máxima probabilidad de θ y luego calcule la estimación para los datos dados.

23. Dos sistemas de computadoras diferentes son supervisados durante un total de n semanas. Sea X_i el número de descomposturas del primer sistema durante la semana i-ésima y suponga que las X_i son independientes y que se extraen de una distribución de Poisson con parámetro μ_1. Asimismo, sea Y_i el número de descomposturas del segundo sistema durante la semana i-ésima y suponga independencia con cada Y_i extraída de una distribución de Poisson con parámetro μ_2. Deduzca los estimadores de máxima probabilidad de μ_1, μ_2 y $\mu_1 - \mu_2$. [*Sugerencia*: utilizando independencia, escriba la función de masa de probabilidad conjunta de las X_i y Y_i juntas.]

24. Un vehículo con un defecto particular en su sistema de control de emisiones es llevado a una serie de mecánicos seleccionados al azar hasta que $r = 3$ de ellos han diagnosticado correctamente el problema. Supongamos que esto requiere de los diagnósticos de 20 mecánicos diferentes (por lo que hubo 17 diagnósticos incorrectos). Sea $p = P$ (diagnóstico correcto), por lo que el estimador de máxima probabilidad es la proporción de todos los mecánicos que bien podría diagnosticar el problema. ¿Cuál el estimador de máxima probabilidad de p? ¿Es el mismo estimador de máxima probabilidad si una muestra aleatoria de 20 mecánicos resulta en tres diagnósticos correctos? Explique. ¿Cómo funciona el estimador de máxima probabilidad en comparación con la estimación resultante de la utilización del estimador imparcial dada en el ejercicio 17?

25. Se determina la resistencia al esfuerzo cortante de soldaduras de 10 puntos de prueba y se obtienen los siguientes datos (lb/pulg^2):

392 376 401 367 389 362 409 415 358 375

a. Suponiendo que la resistencia al esfuerzo cortante está normalmente distribuida, estime la resistencia al esfuerzo cortante promedio verdadera y la desviación estándar de la resistencia al esfuerzo cortante utilizando el método de máxima probabilidad.

b. De nuevo suponiendo una distribución normal, calcule el valor de resistencia por debajo del cual 95% estarán las resistencias de todas las soldaduras. [*Sugerencia*: ¿cuál es el 95º percentil en función de μ y σ? Utilice ahora el principio de invarianza.]

26. Remítase al ejercicio 25. Suponga que decide examinar otra soldadura de puntos de prueba. Sea X = resistencia al esfuerzo

cortante de la soldadura. Use los datos dados para obtener el estimador de máxima probabilidad de $P(X \leq 400)$. [*Sugerencia*: $P(X \leq 400) = \Phi((400 - \mu)/\sigma)$.]

27. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución gamma con parámetros α y β.

a. Deduzca las ecuaciones cuyas soluciones dan los estimadores de máxima probabilidad de α y β. ¿Piensa que pueden ser resueltos explícitamente?

b. Demuestre que el estimador de máxima probabilidad de $\mu = \alpha\beta$ es $\hat{\mu} = \overline{X}$.

28. Si $X_1, X_2, \ldots, X_n$ representan una muestra aleatoria de la distribución de Rayleigh con función de densidad dada en el ejercicio 15, determine:

a. El estimador de máxima probabilidad de θ y luego calcule la estimación para los datos de esfuerzo de vibración dados en ese ejercicio. ¿Es este estimador el mismo estimador insesgado sugerido en el ejercicio 15?

b. El estimador de máxima probabilidad de la mediana de la distribución del esfuerzo de vibración. [*Sugerencia*: exprese primero la mediana en función de θ.]

29. Considere la muestra aleatoria $X_1, X_2, \ldots, X_n$ de la función de densidad de probabilidad exponencial desplazada

$$f(x; \lambda, \theta) = \begin{cases} \lambda e^{-\lambda(x-\theta)} & x \geq \theta \\ 0 & \text{de lo contrario} \end{cases}$$

Con $\theta = 0$ da la función de densidad de probabilidad de la distribución exponencial considerada previamente (con densidad positiva a la derecha de cero). Un ejemplo de la distribución exponencial desplazada apareció en el ejemplo 4.5, en el cual la variable de interés fue el tiempo entre vehículos en el flujo de tránsito, y $\theta = .5$ fue el tiempo entre vehículos mínimo posible.

a. Obtenga los estimadores de máxima probabilidad de θ y λ.

b. Si $n = 10$ observaciones de tiempo entre vehículos son realizadas y se obtienen los siguientes resultados 3.11, .64, 2.55, 2.20, 5.44, 3.42, 10.39, 8.93, 17.82 y 1.30, calcule las estimaciones de θ y λ.

30. En el instante $t = 0$, 20 componentes idénticos son puestos a prueba. La distribución de la vida útil de cada uno es exponencial con parámetro λ. El experimentador deja la instalación de prueba sin supervisar. A su regreso 24 horas más tarde, el experimentador termina de inmediato la prueba después de notar que $y = 15$ de los 20 componentes aún están en operación (así que 5 han fallado). Obtenga el estimador de máxima probabilidad de λ. [*Sugerencia*: sea $Y =$ el número que sobrevive 24 horas. En ese caso $Y \sim \text{Bin}(n, p)$. ¿Cuál es el estimador de máxima probabilidad de p? Observe ahora que $p = P(X_i \geq 24)$, donde X_i está exponencialmente distribuida. Esto relaciona λ con p, de modo que el primero puede ser estimado una vez que lo ha sido el segundo.]

EJERCICIOS SUPLEMENTARIOS (31–38)

31. Se dice que un estimador $\hat{\theta}$ es **consistente** si con cualquier $\epsilon > 0$, $P(|\hat{\theta} - \theta| \geq \epsilon) \to 0$ a medida que $n \to \infty$. Es decir, $\hat{\theta}$ es consistente si, a medida que el tamaño de muestra se hace más grande, es menos y menos probable que $\hat{\theta}$ se aleje más que ϵ del valor verdadero de θ. Demuestre que $\overline{X}$ es un estimador consistente de μ cuando $\sigma^2 < \infty$ mediante la desigualdad de Chebyshev del ejercicio 44 del capítulo 3. [*Sugerencia*: la desigualdad puede ser reescrita en la forma

$$P(|Y - \mu_Y| \geq \epsilon) \leq \sigma_Y^2/\epsilon$$

Ahora identifique Y con $\overline{X}$.]

32. **a.** Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución uniforme en $[0, \theta]$. Entonces el estimador de máxima probabilidad de θ es $\hat{\theta} = Y = \text{máx}(X_i)$. Use el hecho de que $Y \leq y$ si y sólo si cada $X_i \leq y$ para deducir la función de distribución acumulativa de Y. Luego demuestre que la función de densidad de probabilidad de $Y = \text{máx}(X_i)$ es

$$f_Y(y) = \begin{cases} \dfrac{ny^{n-1}}{\theta^n} & 0 \leq y \leq \theta \\ 0 & \text{de lo contrario} \end{cases}$$

b. Use el resultado del inciso (a) para demostrar que el estimador de máxima probabilidad es sesgado pero que $(n + 1)$ $\text{máx}(X_i)/n$ es insesgado.

33. En el instante $t = 0$, hay un individuo vivo en una población. Un **proceso de nacimientos puros** se desarrolla entonces como sigue. El tiempo hasta que ocurre el primer nacimiento está exponencialmente distribuido con parámetro λ. Después del primer nacimiento, hay dos individuos vivos. El tiempo hasta que el primero da a luz otra vez es exponencial con parámetro λ y del mismo modo para el segundo individuo. Por consiguiente, el tiempo hasta el siguiente nacimiento es el mínimo de dos variables (λ) exponenciales, el cual es exponencial con parámetro 2λ. Asimismo, una vez que el segundo nacimiento ha ocurrido, hay tres individuos vivos, de modo que el tiempo hasta el siguiente nacimiento es una variable aleatoria exponencial con parámetro 3λ, y así sucesivamente (aquí se está utilizando la propiedad de no memoria de la distribución exponencial). Suponga que se observa el proceso hasta que el sexto nacimiento ha ocurrido y los tiempos hasta los nacimientos sucesivos son 25.2, 41.7, 51.2, 55.5, 59.5, 61.8 (con los cuales deberá calcular los tiempos entre nacimientos sucesivos). Obtenga el estimador de máxima probabilidad de λ. [*Sugerencia*: la probabilidad es un producto de términos exponenciales.]

34. El **error cuadrático medio** de un estimador $\hat{\theta}$ es $\text{ECM}(\hat{\theta}) = E(\hat{\theta} - \theta)^2$. Si $\hat{\theta}$ es insesgado, entonces $\text{ECM}(\hat{\theta}) = V(\hat{\theta})$, pero en general $\text{ECM}(\hat{\theta}) = V(\hat{\theta}) + (\text{sesgo})^2$. Considere el estimador $\hat{\sigma}^2 = KS^2$, donde $S^2 =$ varianza muestral. ¿Qué valor de K reduce al mínimo el error cuadrático

medio de este estimador cuando la distribución de la población es normal? [*Sugerencia*: se puede demostrar que

$$E[(S^2)^2] = (n + 1)\sigma^4/(n - 1)$$

En general, es difícil determinar $\hat{\theta}$ para reducir al mínimo el ECM($\hat{\theta}$), por lo cual se buscan sólo estimadores insesgados y se reduce al mínimo $V(\hat{\theta})$.]

35. Sean $X_1, \ldots, X_n$ una muestra aleatoria de una función de densidad de probabilidad simétrica con respecto a μ. Un estimador de μ que se ha visto que funciona bien con una amplia variedad de distribuciones subyacentes es el *estimador de Hodges-Lehmann*. Para definirlo, primero calcule para cada $i \leq j$ y cada $j = 1, 2, \ldots, n$ el promedio por pares $\overline{X}_{i,j} = (X_i + X_j)/2$. Entonces el estimador es $\hat{\mu}$ = la mediana de las $\overline{X}_{i,j}$. Calcule el valor de esta estimación con los datos del ejercicio 44 del capítulo 1. [*Sugerencia*: construya una tabla cuadrada con las x_i en el margen izquierdo y en la parte superior. Luego calcule los promedios en la diagonal y encima de ella.]

36. Cuando la distribución de la población es normal, se puede utilizar la mediana estadística $\{|X_1 - \tilde{X}|, \ldots, |X_n - \tilde{X}|\}/.6745$ para estimar σ. Este estimador es más resistente a los efectos de los valores apartados (observaciones alejadas del grueso de los datos) que la desviación estándar muestral. Calcule tanto la estimación puntual correspondiente como s para los datos del ejemplo 6.2.

37. Cuando la desviación estándar muestral S está basada en una muestra aleatoria de una distribución de población normal, se puede demostrar que

$$E(S) = \sqrt{2/(n - 1)}\Gamma(n/2)\sigma/\Gamma((n - 1)/2)$$

Use ésta para obtener un estimador insesgado de σ de la forma cS. ¿Cuál es c cuando $n = 20$?

38. Cada uno de n especímenes tiene que ser pesado dos veces en la misma báscula. Sean X_i y Y_i los dos pesos observados del i-ésimo espécimen. Suponga que X_i y Y_i son independientes uno de otro, cada uno normalmente distribuido con valor medio μ_i (el peso verdadero del espécimen i) y varianza σ^2.

a. Demuestre que el estimador de probabilidad máxima de σ^2 es $\hat{\sigma}^2 = \sum(X_i - Y_i)^2/(4n)$. [*Sugerencia*: si $\bar{z} = (z_1 + z_2)/2$, entonces $\Sigma(z_i - \bar{z})^2 = (z_1 - z_2)^2/2$.]

b. ¿Es el estimador de máxima probabilidad $\hat{\sigma}^2$ un estimador insesgado de σ^2? Determine un estimador insesgado de σ^2. [*Sugerencia*: para cualquier variable aleatoria Z, $E(Z^2) = V(Z) + [E(Z)]^2$. Aplique ésta a $Z = X_i - Y_i$.]

Bibliografía

DeGroot, Morris y Mark Schervish, *Probability and Statistics* (3a. ed.), Addison-Wesley, Boston, MA, 2002. Incluye una excelente discusión tanto de propiedades generales como de métodos de estimación puntual; de particular interés son los ejemplos que muestran cómo los principios y métodos generales pueden dar estimadores insatisfactorios en situaciones particulares.

Devore, Jay y Kenneth Berk, *Modern Mathematical Statistics with Applications*, Thomson-Brooks/Cole, Belmont, CA, 2007. La exposición es un poco más completa y compleja que la de este libro.

Efron, Bradley y Robert Tibshirani, *An Introduction to the Bootstrap*, Chapman and Hall, Nueva York, 1993. La Biblia del bootstrap.

Hoaglin, David, Frederick Mosteller y John Tukey, *Understanding Robust and Exploratory Data Analysis*, Wiley, Nueva York, 1983. Contiene varios buenos capítulos sobre estimación puntual robusta, incluido uno sobre estimación *M*.

Rice, John, *Mathematical Statistics and Data Analysis* (3a. ed.), Thomson-Brooks/Cole, Belmont, CA, 2007. Una agradable mezcla de teoría y datos estadísticos.

7 Intervalos estadísticos basados en una sola muestra

INTRODUCCIÓN

Una estimación puntual, por el hecho de ser un solo número no proporciona información sobre la precisión y confiabilidad de la estimación. Considérese, por ejemplo, utilizar el estadístico $\overline{X}$ para calcular una estimación puntual de la resistencia a la ruptura promedio verdadera (g) de toallas de papel de cierta marca, y supóngase que $\bar{x} = 9322.7$. Debido a la variabilidad del muestreo, virtualmente nunca es el caso de que $\bar{x} = \mu$. La estimación puntual no dice nada sobre qué tan cerca pudiera estar de μ. Una alternativa para reportar un solo valor sensible del parámetro que se está estimando es calcular y reportar un intervalo completo de valores factibles: una *estimación de intervalo* o un *intervalo de confianza* (IC). Un intervalo de confianza siempre se calcula seleccionando primero un *nivel de confianza*, el cual mide el grado de confiabilidad del intervalo. Un intervalo de confianza con 95% de nivel de confianza de la resistencia a la ruptura promedio verdadera podría tener un límite inferior de 9162.5 y un límite superior de 9482.9. Entonces al nivel de confianza de 95%, cualquier valor de μ entre 9162.5 y 9482.5 es factible. Un nivel de confianza de 95% implica que 95% de todas las muestras daría un intervalo que incluye μ o cualquier otro parámetro que se esté estimando, y sólo 5% de las muestras darían un intervalo erróneo. Los niveles de confianza más frecuentemente utilizados son 95%, 99% y 90%. Mientras más alto es el nivel de confianza, más fuerte es la creencia de que el valor del parámetro que se está estimando queda dentro del intervalo (en breve se dará una interpretación de cualquier nivel de confianza particular).

El ancho del intervalo da información sobre la precisión de una estimación de intervalo. Si el nivel de confianza es alto y el intervalo resultante es bastante angosto, el conocimiento del valor del parámetro es razonablemente preciso. Un muy amplio intervalo de confianza, sin embargo, transmite el mensaje de que existe gran canti-

dad de incertidumbre sobre el valor de lo que se está estimando. La figura 7.1 muestra intervalos de confianza de 95% de resistencias a la ruptura promedio verdaderas de dos marcas diferentes de toallas de papel. Uno de estos intervalos sugiere un conocimiento preciso de μ, mientras que el otro sugiere un rango muy amplio de valores factibles.

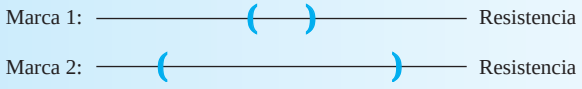


Figura 7.1 Intervalos de confianza que indican información precisa (marca 1) e imprecisa (marca 2) sobre μ

7.1 Propiedades básicas de los intervalos de confianza

Los conceptos y propiedades básicas de los intervalos de confianza son más fáciles de introducir si primero se presta atención a un problema simple, aunque un tanto irreal. Supóngase que el parámetro de interés es una media poblacional μ y que

1. La distribución de la población es normal
2. El valor de la desviación estándar σ de la población es conocido

Con frecuencia es razonable suponer que la distribución de la población es normal. Sin embargo, si el valor de μ es desconocido, no es factible que el valor de σ esté disponible (el conocimiento del centro de una población en general precede a la información con respecto a la dispersión). En las secciones 7.2 y 7.3 se desarrollarán métodos basados en suposiciones menos restrictivas.

Ejemplo 7.1 Ingenieros industriales especialistas en ergonomía se ocupan del diseño de espacios de trabajo y dispositivos operados por trabajadores con objeto de alcanzar una alta productividad y comodidad. El artículo "Studies on Ergonomically Designed Alphanumeric Keyboards" (*Human Factors*, 1985: 175–187) reporta sobre un estudio de altura preferida para un teclado experimental con un gran soporte para el antebrazo y muñeca. Se seleccionó una muestra de $n = 31$ mecanógrafos entrenados y se determinó la altura preferida del teclado de cada mecanógrafo. La altura preferida promedio muestral resultante fue de $\bar{x} = 80.0$ cm. Suponiendo que la altura preferida está normalmente distribuida con $\sigma = 2.0$ cm (un valor sugerido por los datos que aparecen en el artículo), obtenga un intervalo de confianza para μ, la altura promedio verdadera preferida por la población de todos los mecanógrafos experimentados. ■

Se supone que las observaciones muestrales reales $x_1, x_2, \ldots, x_n$ son el resultado de una muestra aleatoria $X_1, \ldots, X_n$ tomada de una distribución normal con valor medio μ y desviación estándar σ. Los resultados del capítulo 5 implican entonces que independientemente del tamaño de muestra n, la media muestral $\bar{X}$ está normalmente distribuida con valor esperado μ y desviación estándar $\sigma/\sqrt{n}$. Si se estandariza $\bar{X}$ restando primero su valor esperado y luego dividiendo entre su desviación estándar se obtiene la variable normal estándar

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \tag{7.1}$$

Como el área bajo la curva normal estándar entre −1.96 y 1.96 es .95,

$$P\left(-1.96 < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < 1.96\right) = .95 \tag{7.2}$$

A continuación manipúlense las desigualdades que están adentro del paréntesis en (7.2) de modo que aparezcan en la forma equivalente $l < \mu < u$, donde los puntos extremos l y u implican $\overline{X}$ y $\sigma/\sqrt{n}$. Esto se logra mediante la siguiente secuencia de operaciones, cada una de las cuales da desigualdades equivalentes a las originales.

1. Multiplíquese por $\sigma/\sqrt{n}$:

$$-1.96 \cdot \frac{\sigma}{\sqrt{n}} < \overline{X} - \mu < 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

2. Réstese $\overline{X}$ de cada término:

$$-\overline{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} < -\mu < -\overline{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

3. Multiplíquese por −1 para eliminar el signo menos enfrente de μ (el cual invierte la dirección de cada desigualdad):

$$\overline{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}} > \mu > \overline{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

es decir,

$$\overline{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

La equivalencia de cada conjunto de desigualdades con el conjunto original implica que

$$P\left(\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = .95 \tag{7.3}$$

El evento en el interior del paréntesis en (7.3) tiene una apariencia poco común; previamente, la cantidad aleatoria aparecía a la mitad con constantes en ambos extremos, como en $a \le Y \le b$. En (7.3) la cantidad aleatoria aparece en los dos extremos, mientras que la constante desconocida μ aparece a la mitad. Para interpretar (7.3), considérese un **intervalo aleatorio** con el punto extremo izquierdo $\overline{X} - 1.96 \cdot \sigma/\sqrt{n}$ y punto extremo derecho $\overline{X} + 1.96 \cdot \sigma/\sqrt{n}$. En notación de intervalo, esto se transforma en

$$\left(\overline{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}},\quad \overline{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right) \tag{7.4}$$

El intervalo (7.4) es aleatorio porque sus dos puntos extremos implican una variable aleatoria. Está centrada en la media muestral $\overline{X}$ y se extiende a $1.96\sigma/\sqrt{n}$ a cada lado de $\overline{X}$. Por consiguiente el ancho del intervalo es $2 \cdot (1.96) \cdot \sigma/\sqrt{n}$, el cual no es aleatorio; sólo la localización del intervalo (su punto medio $\overline{X}$) lo es (figura 7.2). Ahora (7.3) puede ser parafraseado como "*la probabilidad es .95 de que el intervalo aleatorio* (7.4) *incluya o abarque el valor verdadero de* μ". Antes de realizar cualquier experimento y de recolectar cualquier dato, es bastante probable que μ estará dentro del intervalo (7.4).

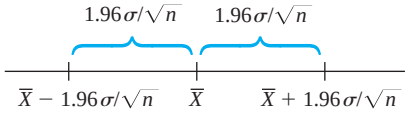


Figura 7.2 Intervalo aleatorio (7.4) con su centro en $\overline{X}$

DEFINICIÓN

Si después de observar $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, se calcula la media muestral observada $\bar{x}$ y luego se sustituye $\bar{x}$ en (7.4) en lugar de $\bar{X}$, el intervalo fijo resultante se llama **intervalo de 95% de confianza para μ.** Este intervalo de confianza se expresa como

$$\left(\bar{x} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right) \text{ es un intervalo de confianza de 95\% para } \mu$$

o como

$$\bar{x} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96 \cdot \frac{\sigma}{\sqrt{n}} \text{ con 95\% de confianza}$$

Una expresión concisa para el intervalo es $\bar{x} \pm 1.96 \cdot \sigma/\sqrt{n}$, donde $-$ da el punto extremo izquierdo (límite inferior) y $+$ da el punto extremo derecho (límite superior).

Ejemplo 7.2 (Continuación del ejemplo 7.1)

Las cantidades requeridas para calcular el intervalo de confianza de 95% para la altura preferida promedio verdadera son $\sigma = 2.0$, $n = 31$ y $\bar{x} = 80.0$. El intervalo resultante es

$$\bar{x} \pm 1.96 \cdot \frac{\sigma}{\sqrt{n}} = 80.0 \pm (1.96)\frac{2.0}{\sqrt{31}} = 80.0 \pm .7 = (79.3, 80.7)$$

Es decir, se puede estar totalmente confiado, en el nivel de confianza de 95%, de que $79.3 < \mu < 80.7$. Este intervalo es relativamente angosto, lo que indica que μ ha sido estimada con bastante precisión. ■

Interpretación de un intervalo de confianza

El nivel de confianza de 95% para el intervalo que se acaba de definir fue heredado del .95 de probabilidad para el intervalo aleatorio (7.4). Los intervalos con otros niveles de confianza serán introducidos en breve. Por ahora, más bien, considérese cómo se puede interpretar el 95% de confianza.

Como se inició con un evento cuya probabilidad era de .95 —que el intervalo aleatorio (7.4) capturaría el valor verdadero de μ—, y luego se utilizaron los datos del ejemplo 7.1 para calcular el intervalo de confianza (79.3, 80.7), es tentador concluir que μ está dentro de este intervalo fijo con probabilidad de .95. Pero al sustituir $\bar{x} = 80.0$ en lugar de $\bar{X}$, toda la aleatoriedad desaparece; el intervalo (79.3, 80.7) no es un intervalo aleatorio y μ es una constante (desafortunadamente desconocida). Es por consiguiente *incorrecto* escribir la proposición $P(\mu$ quede en $(79.3, 80.7)) = .95$.

Una interpretación correcta de "95% de confianza" se basa en la interpretación de probabilidad de frecuencia relativa a largo plazo: decir que un evento A tiene una probabilidad de .95 es decir que si el experimento en el cual se definió A se realiza una y otra vez, a la larga A ocurrirá el 95% del tiempo. Supóngase que se obtiene otra muestra de alturas preferidas por los mecanógrafos y se calcula otro intervalo de 95%. Luego se considera repetir esto con una tercera muestra, una cuarta, una quinta, y así sucesivamente. Sea A el evento en que $\bar{X} - 1.96 \cdot \sigma/\sqrt{n} < \mu < \bar{X} + 1.96 \cdot \sigma/\sqrt{n}$. Ya que $P(A) = .95$, a la larga el 95% de los intervalos de confianza calculados contendrán a μ. Esto se ilustra en la figura 7.3, donde la línea vertical corta el eje de medición en el valor verdadero (pero desconocido) de μ. Observe que 7 de los 100 intervalos mostrados fallan al contener a μ. A la larga, sólo 5% de los intervalos construidos así no contendrán a μ.

De acuerdo con esta interpretación, el nivel de confianza de 95% no es en sí una proposición sobre cualquier intervalo particular tal como (79.3, 80.7). En su lugar pertenece

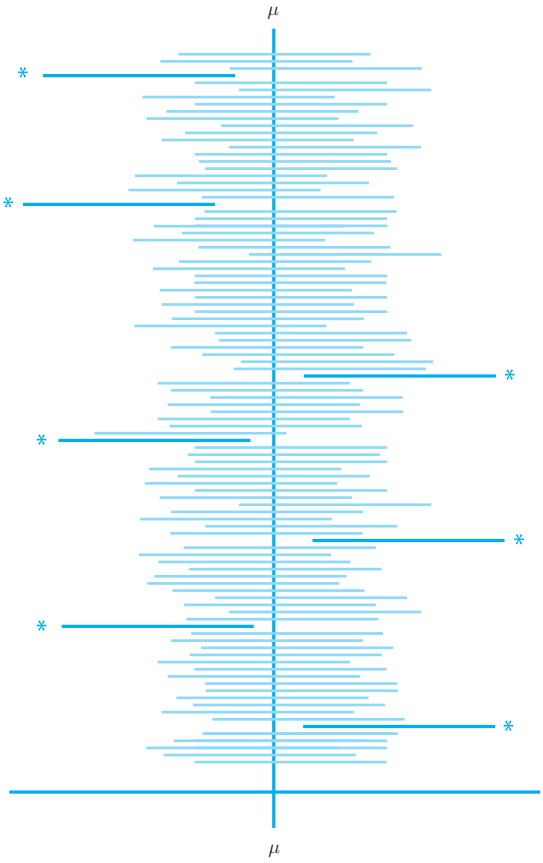


Figura 7.3 Cien niveles de confianza de 95% (los asteriscos identifican intervalos que no incluyen a μ).

a lo que sucedería si se construyera un número muy grande de intervalos parecidos por medio de la misma fórmula de intervalo de confianza. Aunque esto puede parecer no satisfactorio, el origen de la dificultad yace en la interpretación de probabilidad, es válida para una larga secuencia de réplicas de un experimento en lugar de sólo para una. Existe otro método para abordar la construcción e interpretación de intervalos de confianza que utiliza la noción de probabilidad subjetiva y el teorema de probabilidad de Bayes, aunque los detalles técnicos se salen del alcance de este libro; el libro de DeGroot y colaboradores (véase la bibliografía del capítulo 6) es una buena fuente. El intervalo presentado aquí (así como también cada intervalo presentado subsecuentemente) se llama intervalo de confianza "clásico" porque su interpretación se apoya en la noción clásica de probabilidad.

Otros niveles de confianza

El nivel de confianza de 95% fue heredado de la probabilidad de .95 de las desigualdades iniciales que aparecen en (7.2). Si se desea un nivel de confianza de 99%, la probabilidad inicial de .95 debe ser reemplazada por .99, lo que implica cambiar el valor crítico z de 1.96 a 2.58. Un intervalo de confianza de 99% resulta entonces de utilizar 2.58 en lugar de 1.96 en la fórmula para el intervalo de confianza de 95%.

De hecho, cualquier nivel de confianza deseado se obtiene reemplazando 1.96 o 2.58 con el valor crítico normal estándar apropiado. Como la figura 7.4 muestra, utilizando $z_{\alpha/2}$ en lugar de 1.96 se logra una probabilidad de $1 - \alpha$.

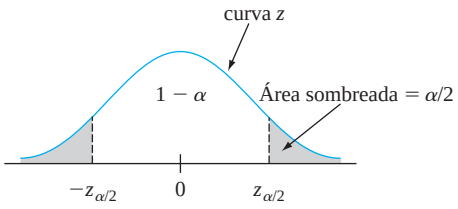


Figura 7.4 $P(-z_{\alpha/2} \le Z < z_{\alpha/2}) = 1 - \alpha$

DEFINICIÓN

La siguiente expresión da un **intervalo de confianza de 100(1 − α)%** para la media μ de una población normal cuando se conoce el valor de σ

$$\left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) \tag{7.5}$$

o, de forma equivalente, por $\bar{x} \pm z_{\alpha/2} \cdot \sigma/\sqrt{n}$.

La fórmula (7.5) para el intervalo de confianza también se puede expresar en palabras como

estimación puntual de $\mu \pm$ (valor crítico z) (error estándar de la media).

Ejemplo 7.3 No hace mucho tiempo que el proceso de producción de una caja de control de un tipo particular para un motor fue modificado. Antes de esta modificación, datos históricos sugirieron que la distribución de los diámetros de agujeros para bujes en las cajas era normal con desviación estándar de .100 mm. Se cree que la modificación no ha afectado la forma de la distribución o la desviación estándar, pero que el valor del diámetro medio pudo haber cambiado. Se selecciona una muestra de 40 cajas y se determina el diámetro de agujero para cada una, y el resultado es un diámetro medio muestral de 5.426 mm. Calcúlese un intervalo de confianza para el diámetro de agujero promedio verdadero utilizando un nivel de confianza de 90%. Esto requiere que $100(1 - \alpha) = 90$, de donde $\alpha = .10$ y $z_{\alpha/2} = z_{.05} = 1.645$ (correspondiente a un área de curva z acumulativa de .9500). El intervalo deseado es entonces

$$5.426 \pm (1.645)\frac{.100}{\sqrt{40}} = 5.426 \pm .026 = (5.400, 5.452)$$

Con un razonablemente alto grado de confianza, se puede decir que $5.400 < \mu < 5.452$. Este intervalo es algo angosto debido a la pequeña cantidad de variabilidad del diámetro del agujero ($\sigma = .100$). ■

Nivel de confianza, precisión y tamaño de muestra

¿Por qué decidirse por un nivel de confianza de 95% cuando un nivel de 99% es alcanzable? Porque el precio pagado por el nivel de confianza más alto es un intervalo más ancho. Como el intervalo de 95% se extiende $1.96 \cdot \sigma/\sqrt{n}$ a cada lado de $\bar{x}$, el ancho del intervalo es $2(1.96) \cdot \sigma/\sqrt{n} = 3.92 \cdot \sigma/\sqrt{n}$. Asimismo, el ancho del intervalo de 99% es $2(2.58) \cdot \sigma/\sqrt{n} = 5.16 \cdot \sigma/\sqrt{n}$. Es decir, se tiene más confianza en el intervalo de 99% precisamente porque es más ancho. Mientras más alto es el grado de confianza deseado, más ancho es el intervalo resultante.

Si se considera que el ancho del intervalo especifica su precisión o exactitud, entonces el nivel de confianza (o confiabilidad) del intervalo está relacionado de manera inversa con su precisión. La estimación de un intervalo altamente confiable puede ser imprecisa

por el hecho de que los puntos extremos del intervalo pueden estar muy alejados, mientras que un intervalo preciso puede acarrear una confiabilidad relativamente baja. Por consiguiente no se puede decir de modo inequívoco que se tiene que preferir un intervalo de 99% a uno de 95%; la ganancia de confiabilidad acarrea una pérdida de precisión.

Una estrategia atractiva es especificar tanto el nivel de confianza deseado como el ancho del intervalo y luego determinar el tamaño de muestra necesario.

Ejemplo 7.4 Un monitoreo exhaustivo de un sistema de tiempo compartido de computadoras sugiere que el tiempo de respuesta a un comando de edición particular está normalmente distribuido con desviación estándar de 25 milisegundos. Se instaló un nuevo sistema operativo y se desea estimar el tiempo de respuesta promedio verdadero μ en el nuevo entorno. Suponiendo que los tiempos de respuesta siguen estando normalmente distribuidos con $\sigma = 25$, ¿qué tamaño de muestra es necesario para asegurarse de que el intervalo de confianza de 95% resultante tiene un ancho de (cuando mucho) 10? El tamaño de muestra n debe satisfacer

$$10 = 2 \cdot (1.96)(25/\sqrt{n})$$

Reordenando esta ecuación se obtiene

$$\sqrt{n} = 2 \cdot (1.96)(25)/10 = 9.80$$

por consiguiente

$$n = (9.80)^2 = 96.04$$

En vista de que n debe ser un entero, se requiere un tamaño de muestra de 97. ■

Una fórmula general para el tamaño de muestra n necesario para garantizar un ancho de intervalo w se obtiene igualando w a $2 \cdot z_{\alpha/2} \cdot \sigma/\sqrt{n}$ y despejando n.

El tamaño de muestra necesario para que el intervalo de confianza (7.5) dé un ancho w es

$$n = \left(2z_{\alpha/2} \cdot \frac{\sigma}{w}\right)^2$$

Mientras más pequeño es el ancho deseado w, más grande debe ser n. Además, n es una función creciente de σ (más variabilidad de la población requiere un tamaño de muestra más grande) y del nivel de confianza $100(1 - \alpha)$ (conforme α decrece, $z_{\alpha/2}$ se incrementa).

La mitad del ancho $1.96\sigma/\sqrt{n}$ del intervalo de confianza de 95% en ocasiones se llama **límite en el error de estimación** asociado con un nivel de confianza de 95%. Es decir, con 95% de confianza, la estimación puntual $\bar{x}$ no estará a más de esta distancia de μ. Antes de obtener datos, es posible que un investigador desee determinar un tamaño de muestra con el cual se logre un valor particular del límite. Por ejemplo, si μ representa la eficiencia de combustible promedio (mpg) de todos los carros de cierto tipo, el objetivo de una investigación puede ser estimar μ dentro de 1 mpg con 95% de confianza. Más generalmente, si se desea estimar μ dentro de una cantidad B (el límite especificado en el error de estimación) con confianza de $100(1 - \alpha)$ %, el tamaño de muestra necesario se obtiene al reemplazar $2/w$ por $1/B$ en el recuadro precedente.

Deducción de un intervalo de confianza

Sean $X_1, X_2, \ldots, X_n$ la muestra en la cual se tiene que basar el intervalo de confianza para un parámetro θ. Supóngase que se puede determinar una variable aleatoria que satisfaga las dos siguientes propiedades:

1. La variable depende funcionalmente tanto de $X_1, \ldots, X_n$ como de θ.
2. La distribución de probabilidad de la variable no depende de θ ni de cualesquiera otros parámetros desconocidos.

Sea $h(X_1, X_2, \ldots, X_n; \theta)$ esta variable aleatoria. Por ejemplo, si la distribución de la población es normal con σ y $\theta = \mu$ conocidos, la variable $h(X_1, \ldots, X_n; \mu) = (\bar{X} - \mu)/(\sigma/\sqrt{n})$ satisface ambas propiedades; claramente depende funcionalmente de μ, no obstante su distribución de probabilidad normal estándar, la cual no depende de μ. En general, la forma de la función h casi siempre se pone de manifiesto al examinar la distribución de un estimador apropiado $\hat{\theta}$.

Con cualquier α entre 0 y 1, se ve que las constantes a y b satisfacen

$$P(a < h(X_1, \ldots, X_n; \theta) < b) = 1 - \alpha \tag{7.6}$$

A causa de la segunda propiedad, a y b no dependen de θ. En el ejemplo normal, $a = -z_{\alpha/2}$ y $b = z_{\alpha/2}$. Ahora supóngase que las desigualdades en (7.6) pueden ser manipuladas para aislar θ, y así obtener la proposición de probabilidad equivalente

$$P(l(X_1, X_2, \ldots, X_n) < \theta < u(X_1, X_2, \ldots, X_n)) = 1 - \alpha$$

Entonces $l(x_1, x_2, \ldots, x_n)$ y $u(x_1, \ldots, x_n)$ son los límites de confianza inferior y superior, respectivamente, para un intervalo de confianza de $100(1 - \alpha)\%$. En el ejemplo normal, se vio que $l(X_1, \ldots, X_n) = \bar{X} - z_{\alpha/2} \cdot \sigma/\sqrt{n}$ y $u(X_1, \ldots, X_n) = \bar{X} + z_{\alpha/2} \cdot \sigma/\sqrt{n}$.

Ejemplo 7.5 Un modelo teórico sugiere que el tiempo hasta la ruptura de un fluido aislante entre electrodos a un voltaje particular tiene una distribución exponencial con parámetro λ (véase la sección 4.4). Una muestra aleatoria de $n = 10$ tiempos de ruptura da los siguientes datos muestrales (en min): $x_1 = 41.53$, $x_2 = 18.73$, $x_3 = 2.99$, $x_4 = 30.34$, $x_5 = 12.33$, $x_6 = 117.52$, $x_7 = 73.02$, $x_8 = 223.63$, $x_9 = 4.00$, $x_{10} = 26.78$. Se desea un intervalo de confianza de 95% para λ y para el tiempo de ruptura promedio verdadero.

Sea $h(X_1, X_2, \ldots, X_n; \lambda) = 2\lambda\sum X_i$. Se puede demostrar que esta variable aleatoria tiene una distribución de probabilidad llamada distribución ji cuadrada con $2n$ grados de libertad (gl) ($\nu = 2n$, donde ν es el parámetro de una distribución ji cuadrada como se menciona en la sección 4.4). La tabla A.7 del apéndice ilustra una curva de densidad ji cuadrada típica y tabula valores críticos que capturan áreas de colas específicas. El número pertinente de grados de libertad en este caso es $2(10) = 20$. La fila $\nu = 20$ de la tabla muestra que 34.170 captura un área de cola superior de .025 y 9.591 captura un área de cola inferior de .025 (área de cola superior de .975). Por consiguiente con $n = 10$,

$$P(9.591 < 2\lambda\sum X_i < 34.170) = .95$$

La división entre $2\sum X_i$ aísla λ y se obtiene

$$P(9.591/(2\sum X_i) < \lambda < (34.170/(2\sum X_i)) = .95$$

El límite inferior del intervalo de confianza de 95% para λ es $9.591/(2\sum x_i)$, y el límite superior es $34.170/(2\sum x_i)$. Con los datos dados $\sum x_i = 550.87$ da el intervalo (.00871, .03101).

El valor esperado de una variable aleatoria exponencial es $\mu = 1/\lambda$. Puesto que

$$P(2\sum X_i/34.170 < 1/\lambda < 2\sum X_i/9.591) = .95$$

el intervalo de confianza de 95% para el tiempo de ruptura promedio verdadero es $(2\sum x_i/34.170, 2\sum x_i/9.591) = (32.24, 114.87)$. Obviamente este intervalo es bastante ancho, lo que refleja una variabilidad sustancial de los tiempos de ruptura y un pequeño tamaño de muestra. ■

En general, los límites de confianza superior e inferior resultan de reemplazar cada $<$ en (7.6) por $=$ y resolviendo para θ. En el ejemplo del fluido aislante que se acaba de considerar, $2\lambda\sum x_i = 34.170$ da $\lambda = 34.170/(2\sum x_i)$ como límite de confianza superior y el límite inferior se obtiene con la otra ecuación. Obsérvese que los dos límites de intervalo no están equidistantes de la estimación puntual, en vista de que el intervalo no es de la forma $\hat{\theta} \pm c$.

Intervalos de confianza bootstrap

La técnica bootstrap se introdujo en el capítulo 6 como una forma de estimar $\sigma_{\hat{\theta}}$. También puede ser aplicada para obtener un intervalo de confianza para θ. Considérese de nuevo la estimación de la media μ de una distribución normal cuando σ es conocido. Reemplácese μ con θ y úsese $\hat{\theta} = \overline{X}$ como estimador puntual. Obsérvese que $1.96\sigma/\sqrt{n}$ es el 97.5º percentil de la distribución de $\hat{\theta} - \theta$ [esto es, $P(\overline{X} - \mu < 1.96\sigma/\sqrt{n}) = P(Z < 1.96) = .9750$]. Del mismo modo, $-1.96\sigma/\sqrt{n}$ es el 2.5º percentil, por consiguiente

$$\begin{aligned} .95 &= P(2.5^{\text{o}} \text{ percentil} < \hat{\theta} - \theta < 97.5^{\text{o}} \text{ percentil}) \\ &= P(\hat{\theta} - 2.5^{\text{o}} \text{ percentil} > \theta > \hat{\theta} - 97.5^{\text{o}} \text{ percentil}) \end{aligned}$$

Es decir, con

$$\begin{aligned} l &= \hat{\theta} - 97.5^{\text{o}} \text{ percentil de } \hat{\theta} - \theta \\ u &= \hat{\theta} - 2.5^{\text{o}} \text{ percentil de } \hat{\theta} - \theta \end{aligned} \tag{7.7}$$

el intervalo de confianza para θ es (l, u). En muchos casos, los percentiles en (7.7) no pueden ser calculados, pero sí *pueden* ser estimados con muestras bootstrap. Supóngase que se obtienen $B = 1000$ muestras bootstrap y se calculan $\hat{\theta}^*_1, \ldots, \hat{\theta}^*_{1000}$ y $\bar{\theta}^*$ seguidos por las 1000 diferencias $\hat{\theta}^*_1 - \bar{\theta}^*, \ldots, \hat{\theta}^*_{1000} - \bar{\theta}^*$. La 25º más grande y la 25º más pequeña de estas diferencias son estimaciones de los percentiles desconocidos en (7.7). Consúltese los libros de Devore y Berk o de Efron citados en el capítulo 6 para más información.

EJERCICIOS Sección 7.1 (1–11)

1. Considere una distribución de población normal con el valor de σ conocido.
- **a.** ¿Cuál es el nivel de confianza para el intervalo $\bar{x} \pm 2.81\sigma/\sqrt{n}$?
- **b.** ¿Cuál es el nivel de confianza para el intervalo $\bar{x} \pm 1.44\sigma/\sqrt{n}$?
- **c.** ¿Qué valor de $z_{\alpha/2}$ en la fórmula de intervalo de confianza (7.5) da un nivel de confianza de 99.7%?
- **d.** Responda la pregunta hecha en el inciso (c) para un nivel de confianza de 75%.

2. Cada uno de los siguientes es un intervalo de confianza para μ = frecuencia de resonancia promedio (es decir, media de la población) verdadera (Hz) para todas las raquetas de tenis de un tipo:

(114.4, 115.6) (114.1, 115.9)

- **a.** ¿Cuál es el valor de la frecuencia de resonancia media muestral?
- **b.** Ambos intervalos se calcularon con los mismos datos muestrales. El nivel de confianza para uno de estos intervalos es de 90% y para el otro es de 99%. ¿Cuál de los intervalos tiene el nivel de confianza de 90% y por qué?

3. Suponga que se selecciona una muestra aleatoria de 50 botellas de una marca particular de jarabe para la tos y se determina el contenido de alcohol de cada una. Sea μ el contenido promedio de alcohol de la población de todas las botellas de la marca estudiada. Suponga que el intervalo de confianza de 95% resultante es (7.8, 9.4).
- **a.** ¿Un intervalo de confianza de 90% calculado con esta muestra habría resultado más angosto o más ancho que el intervalo dado? Explique su razonamiento.
- **b.** Considere la siguiente proposición: existe 95% de probabilidades de que μ esté entre 7.8 y 9.4. ¿Es correcta esta proposición? ¿ Por qué sí o por qué no?
- **c.** Considere la siguiente proposición: se puede estar totalmente confiado de que 95% de todas las botellas de este tipo de jarabe para la tos tienen un contenido de alcohol de entre 7.8 y 9.4. ¿Es correcta esta proposición? ¿Por qué sí o por qué no?
- **d.** Considere la siguiente proposición: si el proceso de selección de una muestra de tamaño 50 y el cálculo del intervalo de 95% correspondiente se repite 100 veces, 95 de los intervalos resultantes incluirán μ. ¿Es correcta esta proposición? ¿Por qué sí o por qué no?

4. Se desea un intervalo de confianza para la pérdida por carga parásita promedio verdadera μ (watts) de cierto tipo de motor de inducción cuando la corriente a través de la línea se mantiene en 10 amps a una velocidad de 1500 rpm. Suponga que la pérdida por carga parásita está normalmente distribuida con $\sigma = 3.0$.
- **a.** Calcule un intervalo de confianza de 95% para μ cuando $n = 25$ y $\bar{x} = 58.3$.
- **b.** Calcule un intervalo de confianza de 95% para μ cuando $n = 100$ y $\bar{x} = 58.3$.
- **c.** Calcule un intervalo de confianza de 99% para μ cuando $n = 100$ y $\bar{x} = 58.3$.
- **d.** Calcule un intervalo de confianza de 82% para μ cuando $n = 100$ y $\bar{x} = 58.3$.
- **e.** ¿Qué tan grande debe ser n si el ancho del intervalo de 99% para μ tiene que ser 1.0?

5. Suponga que la porosidad al helio (en porcentaje) de muestras de carbón tomadas de cualquier costura particular está normalmente distribuida con desviación estándar verdadera de .75.
- **a.** Calcule un intervalo de confianza de 95% para la porosidad promedio verdadera de una costura si la porosidad promedio en 20 especímenes de la costura fue de 4.85.
- **b.** Calcule un intervalo de confianza de 98% para la porosidad promedio verdadera de otra costura basada en 16 especímenes con porosidad promedio muestral de 4.56.
- **c.** ¿Qué tan grande debe ser un tamaño de muestra si el ancho del intervalo de 95% tiene que ser de .40?
- **d.** ¿Qué tamaño de muestra se necesita para estimar la porosidad promedio verdadera dentro de .2 con confianza de 99%?

6. Con base en pruebas extensas, se sabe que el punto de cedencia de un tipo particular de varilla de refuerzo de acero suave está normalmente distribuido con $\sigma = 100$. La composición de la varilla se modificó un poco, pero no se cree que la alteración haya afectado la normalidad o el valor de σ.
- **a.** Suponiendo que éste sea el caso, si una muestra de 25 varillas modificadas dio por resultado un punto de cedencia promedio muestral de 8439 lb, calcule un intervalo de confianza de 90% para el punto de cedencia promedio verdadero de la varilla modificada.
- **b.** ¿Cómo modificaría el intervalo del inciso (a) para obtener un nivel de confianza de 92%?

7. ¿En cuánto se debe incrementar el tamaño de muestra n si el ancho del intervalo de confianza (7.5) tiene que ser reducido a la mitad? Si el tamaño de muestra se incrementa por un factor de 25, ¿qué efecto tendrá esto en el ancho del intervalo? Justifique sus aseveraciones.

8. Sea $\alpha_1 > 0$, $\alpha_2 > 0$, con $\alpha_1 + \alpha_2 = \alpha$. Entonces

$$P\left(-z_{\alpha_1} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha_2}\right) = 1 - \alpha$$

- **a.** Use esta ecuación para obtener una expresión más general para un intervalo de confianza de $100(1 - \alpha)\%$ para μ del cual el intervalo (7.5) es un caso especial.
- **b.** Sea $\alpha = .05$ y $\alpha_1 = \alpha/4$, $\alpha_2 = 3\alpha/4$. ¿Esto resulta en un intervalo más angosto o más ancho que el intervalo (7.5)?

9. **a.** En las mismas condiciones que aquellas que conducen al intervalo (7.5), $P[(\bar{X} - \mu)/(\sigma/\sqrt{n}) < 1.645] = .95$. Use esta expresión para deducir un intervalo unilateral para μ de ancho infinito y que proporcione un límite de confianza inferior para μ. ¿Cuál es el intervalo para los datos del ejercicio 5(a)?
- **b.** Generalice el resultado del inciso (a) para obtener un límite inferior con nivel de confianza de $100(1 - \alpha)\%$.
- **c.** ¿Cuál es un intervalo análogo al del inciso (b) que proporcione un límite superior para μ? Calcule este intervalo de 99% para los datos del ejercicio 4(a).

10. Una muestra aleatoria de $n = 15$ bombas térmicas de cierto tipo produjo las siguientes observaciones de vida útil (en años):

2.0	1.3	6.0	1.9	5.1	.4	1.0	5.3
15.7	.7	4.8	.9	12.2	5.3	.6	

- **a.** Suponga que la distribución de la vida útil es exponencial y use un argumento paralelo al del ejemplo 7.5 para obtener un intervalo de confianza de 95% para la vida útil esperada (promedio verdadero).
- **b.** ¿Cómo debería modificarse el intervalo del inciso (a) para obtener un nivel de confianza de 99%?
- **c.** ¿Cuál es un intervalo de confianza de 95% para la desviación estándar de la distribución de la vida útil? [*Sugerencia*: ¿cuál es la desviación estándar de una variable aleatoria exponencial?]

11. Considere los siguientes 1000 intervalos de confianza de 95% para μ que un consultor estadístico obtendrá para varios clientes. Suponga que se seleccionan independientemente uno de otro los conjuntos de datos en los cuales están basados los intervalos. ¿Cuántos de estos 1000 intervalos espera que capturen el valor correspondiente de μ? ¿Cuál es la probabilidad de que entre 940 y 960 de estos intervalos contengan el valor correspondiente de μ? [*Sugerencia*: sea Y = el número entre los 1000 intervalos que contienen a μ. ¿Qué clase de variable aleatoria es Y?]

7.2 Intervalos de confianza de muestra grande para una media y proporción de población

Se supuso en el intervalo de confianza para μ dado en la sección previa que la distribución de la población es normal con el valor de σ conocido. A continuación se presenta un intervalo de confianza de muestra grande cuya validez no requiere estas suposiciones. Después de demostrar cómo el argumento que lleva a este intervalo se aplica en forma extensa para producir otros intervalos de muestra grande, habrá que enfocarse en un intervalo para una proporción de población p.

Intervalo de muestra grande para μ

Sean $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población con media μ y desviación estándar σ. Siempre que n es grande, el teorema del límite central implica que $\overline{X}$ tiene de manera aproximada una distribución normal cualquiera que sea la naturaleza de la distribución de la población. Se deduce entonces que $Z = (\overline{X} - \mu)/(\sigma/\sqrt{n})$ tiene aproximadamente una distribución estándar normal, de modo que

$$P\left(-z_{\alpha/2} < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) \approx 1 - \alpha$$

Un argumento paralelo al dado en la sección 7.1 da $\bar{x} \pm z_{\alpha/2} \cdot \sigma/\sqrt{n}$ como intervalo de confianza de muestra grande para μ con un nivel de confianza de *aproximadamente* $100(1 - \alpha)\%$. Es decir, cuando n es grande, el intervalo de confianza para μ dado antes permanece válido cualquiera que sea la distribución de la población, siempre que el calificador esté insertado "aproximadamente" enfrente del nivel de confianza.

Una dificultad práctica con este desarrollo es que el cálculo del intervalo de confianza requiere el valor de σ, el cual rara vez es conocido. Considérese la variable estandarizada $(\overline{X} - \mu)/(S/\sqrt{n})$, en la cual la desviación estándar muestral S ha reemplazado a σ. Previamente había aleatoriedad sólo en el numerador de Z gracias a $\overline{X}$. En la nueva variable estandarizada, tanto $\overline{X}$ *como* S cambian de valor de una muestra a otra. Así que aparentemente la distribución de la nueva variable deberá estar más dispersa que la curva z para reflejar la variación extra en el denominador. Esto en realidad es cierto cuando n es pequeño. Sin embargo, con n grande la sustitución de S en lugar de σ agrega un poco de variabilidad extra, así que esta variable también tiene aproximadamente una distribución normal estándar. La manipulación de la variable en la proposición de probabilidad, como en el caso de σ conocida, da un intervalo de confianza de muestra grande general para μ.

PROPOSICIÓN

Si n es suficientemente grande, la variable estandarizada

$$Z = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

tiene aproximadamente una distribución normal estándar. Esto implica que

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \tag{7.8}$$

es un **intervalo de confianza de muestra grande para μ** con nivel de confianza aproximadamente de $100(1 - \alpha)\%$. Esta fórmula es válida sin importar la forma de la distribución de la población.

Es decir, el intervalo de confianza (7.8) es la

estimación puntual de $\mu \pm$ (z valor crítico) (error estándar estimado de la media).

En general, $n > 40$ será suficiente para justificar el uso de este intervalo. Esto es algo más conservador que la regla empírica del teorema del límite central debido a la variabilidad adicional introducida por el uso de S en lugar de σ.

Ejemplo 7.6 ¿Siempre quiso tener un Porsche? El autor pensó que tal vez podía permitirse un Boxster, el modelo más barato. Así que se fue a www.cars.com el 18 de noviembre de 2009 y encontró un total de 1113 automóviles de este tipo en la lista. Preguntando, los precios iban desde $ 3499 a $ 130,000 (el precio de este último fue uno de los dos que excedían los

$ 70,000). Los precios lo deprimieron, por lo que en cambio se centró en las lecturas del odómetro (millas). Aquí se presentan las lecturas de una muestra de 50 de estos Boxster:

2948	2996	7197	8338	8500	8759	12710	12925
15767	20000	23247	24863	26000	26210	30552	30600
35700	36466	40316	40596	41021	41234	43000	44607
45000	45027	45442	46963	47978	49518	52000	53334
54208	56062	57000	57365	60020	60265	60803	62851
64404	72140	74594	79308	79500	80000	80000	84000
113000	118634						

Una gráfica de caja de los datos (figura 7.5) muestra que, a excepción de los dos valores límite en el extremo superior, la distribución de valores es bastante simétrica (de hecho, una gráfica de probabilidad normal muestra un patrón bastante lineal, aunque los puntos correspondientes a las dos observaciones más pequeñas y a las dos mayores están un tanto alejadas de un ajuste lineal a través de los puntos restantes).

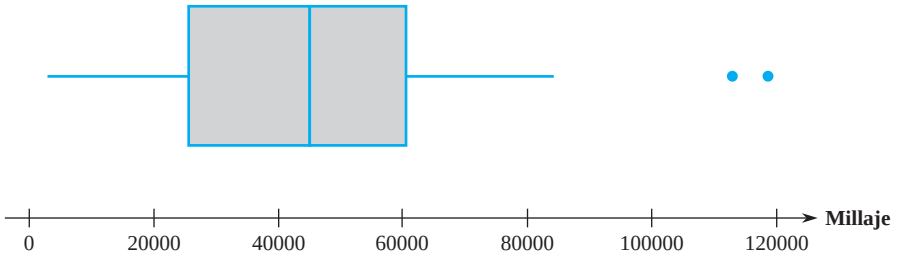


Figura 7.5 Diagrama de caja para las lecturas del odómetro del ejemplo 7.6

Las cantidades resumidas incluyen $n = 50$, $\bar{x} = 45{,}679.4$, $\tilde{x} = 45{,}013.5$, $s = 26{,}641.675$, $f_s = 34{,}265$. La media y la mediana están relativamente cerca (si los dos valores mayores fueran reducidos por 30,000, la media bajaría a 44,479.4, mientras que la mediana no se vería afectada). El diagrama de caja y las magnitudes de s y f_s respecto a la media y la mediana de ambos indican una cantidad considerable de variabilidad. El intervalo de confianza de 95% requiere que $z_{.025} = 1.96$, y el intervalo es entonces

$$45{,}679.4 \pm (1.96)\left(\frac{26{,}641.675}{\sqrt{50}}\right) = 45{,}679.4 \pm 7384.7$$
$$= (38{,}294.7,\ 53{,}064.1)$$

Es decir, $38{,}294.7 < \mu < 53{,}064.1$ con un nivel de confianza de aproximadamente 95%. Este intervalo es bastante amplio debido a un tamaño de muestra de 50, que aunque es grande por nuestra regla general, no es lo suficientemente grande como para superar la variabilidad en la muestra. No tenemos una estimación muy precisa de la población media de la lectura del odómetro.

¿Es el intervalo que hemos calculado uno de los 95% que en el largo plazo incluyen el parámetro calculado o es uno de los "malos" del 5% que no lo hace? Sin saber el valor de μ, no podemos decir. Recuerde que el nivel de confianza se refiere al porcentaje de captura a largo plazo cuando la fórmula se utiliza repetidamente en varias muestras; no se puede interpretar para una sola muestra y el intervalo resultante. ■

Desafortunadamente, la selección del tamaño de muestra para que dé un ancho de intervalo deseado no es simple en este caso como lo fue en el caso de σ conocida. Por eso el ancho de (7.8) es $2z_{\alpha/2}s/\sqrt{n}$. Como el valor de s no está disponible antes de que los datos

hayan sido recopilados, el ancho del intervalo no puede ser determinado tan sólo con la selección de n. La única opción para un investigador que desea especificar un ancho deseado es hacer una suposición educada de cuál podría ser el valor de s. Siendo conservador y suponiendo un valor más grande de s, se seleccionará un n más grande de lo necesario. El investigador puede ser capaz de especificar un valor razonablemente preciso del rango de población (la diferencia entre los valores más grande y más pequeño). Entonces si la distribución de la población no es demasiado asimétrica, si se divide el rango entre 4 se obtiene un valor aproximado de lo que s podría ser.

Ejemplo 7.7 El tiempo de carga (minutos) para el acero de carbono en un tipo de horno de hogar abierto se determinará para cada calor en una muestra de tamaño n. Si el investigador cree que casi todos los tiempos en la distribución están entre 320 y 440, ¿qué tamaño de la muestra sería apropiado para estimar el tiempo promedio real a cuando mucho 5 minutos con un nivel de confianza del 95%?

Un valor razonable para s es $(440 - 320)/4 = 30$. Por tanto

$$n = \left[\frac{(1.96)(30)}{5}\right]^2 = 138.3$$

Dado que el tamaño de la muestra debe ser un número entero, $n = 139$ debe ser utilizado. Tenga en cuenta que la estimación está dentro de 5 minutos con el nivel de confianza especificado que es equivalente a un ancho de intervalo de confianza de 10 minutos. ■

Un intervalo de confianza de muestra grande general

Los intervalos de muestra grande $\bar{x} \pm z_{\alpha/2} \cdot \sigma/\sqrt{n}$ y $\bar{x} \pm z_{\alpha/2} \cdot s/\sqrt{n}$ son casos especiales de un intervalo de confianza de muestra grande general para un parámetro θ. Suponga que $\hat{\theta}$ es un estimador que satisface las siguientes propiedades: (1) Tiene aproximadamente una distribución normal; (2) es insesgado (por lo menos aproximadamente); y (3) una expresión para $\sigma_{\hat{\theta}}$, la desviación estándar de $\hat{\theta}$, está disponible. Por ejemplo, en el caso $\theta = \mu$, $\hat{\mu} = \bar{X}$ es un estimador insesgado cuya distribución es aproximadamente normal cuando n es grande y $\sigma_{\hat{\mu}} = \sigma_{\bar{X}} = \sigma/\sqrt{n}$. Estandarizando $\hat{\theta}$ se obtiene la variable aleatoria $Z = (\hat{\theta} - \theta)/\sigma_{\hat{\theta}}$, la cual tiene aproximadamente una distribución normal estándar. Esto justifica la proposición de probabilidad

$$P\left(-z_{\alpha/2} < \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} < z_{\alpha/2}\right) \approx 1 - \alpha \qquad (7.9)$$

Suponga, primero, que $\sigma_{\hat{\theta}}$ no involucra ningún parámetro desconocido (p. ej., σ conocida en el caso $\theta = \mu$). Entonces si se reemplaza cada $<$ en (7.9) por $=$ se obtiene $\theta = \hat{\theta} \pm z_{\alpha/2} \cdot \sigma_{\hat{\theta}}$, por consiguiente los límites de confianza inferior y superior son $\hat{\theta} - z_{\alpha/2} \cdot \sigma_{\hat{\theta}}$ y $\hat{\theta} + z_{\alpha/2} \cdot \sigma_{\hat{\theta}}$, respectivamente. Suponga ahora que $\sigma_{\hat{\theta}}$ no implica a θ pero sí implica por lo menos otro parámetro desconocido. Sea $s_{\hat{\theta}}$ la estimación de $\sigma_{\hat{\theta}}$ obtenida utilizando estimaciones en lugar de los parámetros desconocidos (p. ej., $s/\sqrt{n}$ estima $\sigma/\sqrt{n}$). En condiciones generales (esencialmente que $s_{\hat{\theta}}$ se aproxime a $\sigma_{\hat{\theta}}$ con la mayoría de las muestras), un intervalo de confianza válido es $\hat{\theta} \pm z_{\alpha/2} \cdot s_{\hat{\theta}}$. El intervalo muestral grande $\bar{x} \pm z_{\alpha/2} \cdot s/\sqrt{n}$ es un ejemplo.

Por último, suponga que $\sigma_{\hat{\theta}}$ implica el θ desconocido. Éste es el caso, por ejemplo, cuando $\theta = p$, una proporción de la población. Entonces $(\hat{\theta} - \theta)/\sigma_{\hat{\theta}} = z_{\alpha/2}$ puede ser difícil de resolver. Con frecuencia se puede obtener una solución aproximada reemplazando θ en $\sigma_{\hat{\theta}}$ por su estimación $\hat{\theta}$. Esto da una desviación estándar estimada $s_{\hat{\theta}}$ y el intervalo correspondiente es de nuevo $\hat{\theta} \pm z_{\alpha/2} \cdot s_{\hat{\theta}}$.

Es decir, este intervalo de confianza es una

estimación puntual de $\theta \pm$ (valor crítico z) (error estándar estimado del estimador).

Un intervalo de confianza para una proporción de población

Sea p la proporción de "éxitos" en una población, donde *éxito* identifica a un individuo u objeto que tiene una propiedad específica (p. ej., individuos que se graduaron en una universidad, computadoras que no requieren servicio de garantía, etc.). Una muestra aleatoria de n individuos tiene que ser seleccionada y X es el número de éxitos en la muestra. Siempre que n sea pequeño comparado con el tamaño de la población, X puede ser considerada como una variable aleatoria binomial con $E(X) = np$ y $\sigma_X = \sqrt{np(1-p)}$. Además, si tanto $np \geq 10$ como $nq \geq 10$, $(q = 1 - p)$, X tiene aproximadamente una distribución normal.

El estimador natural de p es $\hat{p} = X/n$, la fracción muestral de éxitos. Como $\hat{p}$ es simplemente X multiplicada por la constante $1/n$, $\hat{p}$ también tiene aproximadamente una distribución normal. Como se muestra en la sección 6.1, $E(\hat{p}) = p$ (insesgado) y $\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$. La desviación estándar $\sigma_{\hat{p}}$ implica el parámetro desconocido p. Si se estandariza $\hat{p}$ restando p y dividiendo entre $\sigma_{\hat{p}}$ entonces se tiene

$$P\left(-z_{\alpha/2} < \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} < z_{\alpha/2}\right) \approx 1 - \alpha$$

Procediendo como se sugirió en la subsección "Deducción de un intervalo de confianza" (sección 7.1), los límites de confianza se obtienen al reemplazar cada $<$ por $=$ y resolver la ecuación cuadrática resultante para p. Esto da las dos raíces

$$p = \frac{\hat{p} + z_{\alpha/2}^2/2n}{1 + z_{\alpha/2}^2/n} \pm z_{\alpha/2}\frac{\sqrt{\hat{p}(1-\hat{p})/n + z_{\alpha/2}^2/4n^2}}{1 + z_{\alpha/2}^2/n}$$

$$= \tilde{p} \pm z_{\alpha/2}\frac{\sqrt{\hat{p}(1-\hat{p})/n + z_{\alpha/2}^2/4n^2}}{1 + z_{\alpha/2}^2/n}$$

PROPOSICIÓN

Sea $\tilde{p} = \dfrac{\hat{p} + z_{\alpha/2}^2/2n}{1 + z_{\alpha/2}^2/n}$. Entonces, un **intervalo de confianza para una proporción de población p** con nivel de confianza aproximadamente de $100(1-\alpha)\%$ tiene

$$\tilde{p} \pm z_{\alpha/2}\frac{\sqrt{\hat{p}\hat{q}/n + z_{\alpha/2}^2/4n^2}}{1 + z_{\alpha/2}^2/n} \qquad (7.10)$$

donde $\hat{q} = 1 - \hat{p}$ y como antes, el signo $(-)$ en la ecuación 7.10 corresponde al límite de confianza inferior y el signo $(+)$ al límite de confianza superior.

Esto se denomina a menudo como la *puntuación del intervalo de confianza* para p.

Si el tamaño n de la muestra es bastante grande, entonces $z^2/2n$ suele ser insignificante (pequeño) comparado con $\hat{p}$ y z^2/n es insignificante comparado con 1, partiendo de que $\tilde{p} \approx \hat{p}$. En este caso $z^2/4n^2$ también es despreciable comparado con pq/n (n^2 es un divisor mucho más grande que n); como resultado, el término dominante en la expresión $\pm$ es $z_{\alpha/2}$ es $\sqrt{\hat{p}\hat{q}/n}$ y el intervalo de puntuación es aproximadamente

$$\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n} \qquad (7.11)$$

Este último intervalo tiene la forma general $\hat{\theta} \pm z_{\alpha/2}\hat{\sigma}_{\hat{\theta}}$ de un amplio intervalo de la muestra sugerido en la última subsección. La aproximación del intervalo de confianza (7.11) es el que durante décadas ha aparecido en libros de texto de introducción a la estadística. Está

claro que tiene una forma mucho más simple y más atractiva que la puntuación del intervalo de confianza. Así que, ¿por qué molestarse con este último?

Primero que todo, supóngase que se utiliza $z_{.025} = 1.96$ en la fórmula tradicional (7.11). Entonces, nuestro nivel de confianza *nominal* (el que creo que va a comprar utilizando este valor crítico z) es de aproximadamente 95%. Así que antes de seleccionar una muestra, la probabilidad de que el intervalo aleatorio incluya el valor real de p (es decir, la *probabilidad de cobertura*) debe ser de .95. Pero, como muestra la figura 7.6 para el caso $n = 100$, la probabilidad de cobertura real de este intervalo puede variar considerablemente de la probabilidad nominal de .95, en particular cuando p no está cerca de .5 (la gráfica de probabilidad de cobertura frente a p es muy irregular debido a que la distribución subyacente de probabilidad binomial es discreta y no continua). Esto es en general una deficiencia del intervalo tradicional, el nivel de confianza real puede ser bastante diferente del nivel nominal, incluso para tamaños de muestra razonablemente grandes. Investigaciones recientes han demostrado que el intervalo de la puntuación rectifica este comportamiento para prácticamente todos los tamaños de las muestras y los valores de p, su nivel de confianza real será bastante cercano al nivel nominal especificado por la elección de $z_{\alpha/2}$. Esto se debe en gran parte al hecho de que el intervalo de la puntuación se desplaza un poco hacia el .5 en comparación con los intervalos tradicionales. En particular, el punto medio $\tilde{p}$ del intervalo de la puntuación es siempre un poco más cercano a .5 que el punto medio $\hat{p}$ del intervalo tradicional. Esto es especialmente importante cuando p está cerca de 0 o 1.

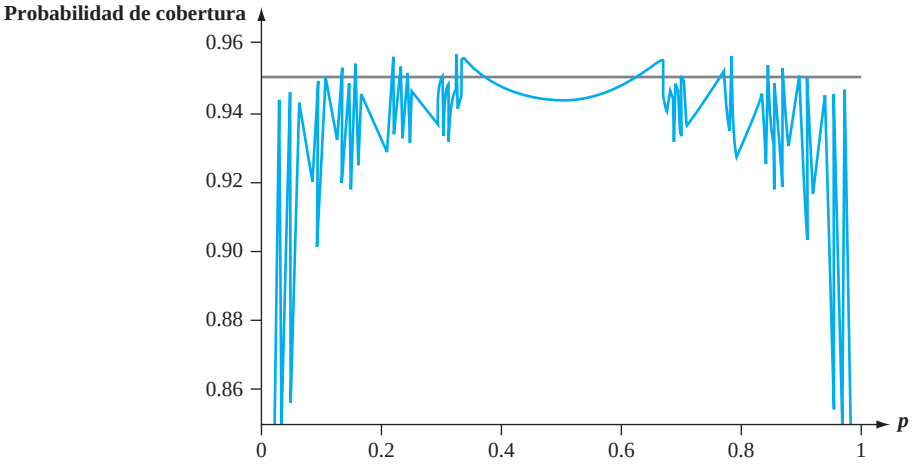


Figura 7.6 Probabilidad de cobertura real para el intervalo (7.11) para variaciones en los valores de p cuando $n = 100$

Además, el intervalo de la puntuación se puede utilizar con casi todos los tamaños de muestra y valores de los parámetros. No es necesario controlar las condiciones $n\hat{p} \geq 10$ y $n(1 - \hat{p}) \geq 10$ que se requerirían al emplear intervalos tradicionales. Así que en lugar de preguntar cuándo n es suficientemente grande para (7.11) para obtener una buena aproximación a (7.10), nuestra recomendación es que la puntuación del intervalo de confianza debe usarse *siempre*. El leve aburrimiento adicional de los cálculos se ve compensado por las propiedades deseables del intervalo.

Ejemplo 7.8 El artículo "Repeatability and Reproducibility for Pass/Fail Data" (*J. of Testing and Eval.*, 1997: 151–153) reportó que en $n = 48$ ensayos en un laboratorio particular, 16 dieron por resultado la ignición de un tipo particular de sustrato por un cigarrillo encendido. Sea p la proporción a largo plazo de todos los ensayos que producirían ignición. Una estimación

puntual de p es $\hat{p} = 16/48 = .333$. Un intervalo de confianza para p con un nivel de confianza de aproximadamente 95% es

$$\frac{.333 + (1.96)^2/96}{1 + (1.96)^2/48} \pm (1.96)\frac{\sqrt{(.333)(.667)/48 + (1.96)^2/9216}}{1 + (1.96)^2/48}$$
$$= .345 \pm .129 = (.216, .474)$$

Este intervalo es bastante amplio ya que un tamaño de muestra de 48 no es tan grande al estimar una proporción.

El intervalo tradicional es

$$.333 \pm 1.96\sqrt{(.333)(.667)/48} = .333 \pm .133 = (.200, .466)$$

Estos dos intervalos concordarían mucho más si el tamaño de muestra fuera sustancialmente más grande. ■

Si se iguala el ancho del intervalo de confianza para p con el ancho preespecificado w se obtiene una ecuación cuadrática para el tamaño de muestra n necesario para dar un intervalo con un grado de precisión deseado. Si se suprime el subíndice en $z_{\alpha/2}$, la solución es

$$n = \frac{2z^2\hat{p}\hat{q} - z^2w^2 \pm \sqrt{4z^4\hat{p}\hat{q}(\hat{p}\hat{q} - w^2) + w^2z^4}}{w^2} \qquad (7.12)$$

Omitiendo los términos en el numerador que implican w^2 se obtiene

$$n \approx \frac{4z^2\hat{p}\hat{q}}{w^2}$$

Esta última expresión es lo que resulta de igualar el ancho del intervalo tradicional con w.

Estas fórmulas desafortunadamente implican la $\hat{p}$ desconocida. El método más conservador es aprovechar el hecho de que $\hat{p}\hat{q}[= \hat{p}(1 - \hat{p})]$ es un máximo cuando $\hat{p} = .5$. Por consiguiente, si se utiliza $\hat{p} = \hat{q} = .5$ en (7.12), el ancho será cuando mucho w haciendo caso omiso de qué valor de $\hat{p}$ resulte de la muestra. De manera alternativa, si el investigador cree de manera firme, basado en información previa, que $p \leq p_0 \leq .5$, en ese caso se utiliza p_0 en lugar de $\hat{p}$. Un comentario similar es válido cuando $p \geq p_0 \geq .5$.

Ejemplo 7.9 El ancho del intervalo de confianza de 95% en el ejemplo 7.8 es .258. El valor de n necesario para garantizar un ancho de .10 independientemente del valor de $\hat{p}$ es

$$n = \frac{2(1.96)^2(.25) - (1.96)^2(.01) \pm \sqrt{4(1.96)^4(.25)(.25 - .01) + (.01)(1.96)^4}}{.01} = 380.3$$

Por consiguiente se deberá utilizar un tamaño de muestra de 381. La expresión para n basada en el intervalo de confianza tradicional da un valor un poco más grande que 385. ■

Intervalos de confianza unilaterales (límites de confianza)

Los intervalos de confianza discutidos hasta ahora dan tanto un límite de confianza inferior *como* uno superior para el parámetro que se está estimando. En algunas circunstancias, es posible que un investigador desee sólo uno de estos dos tipos de límites. Por ejemplo, es posible que un psicólogo desee calcular un límite de confianza superior de 95% para el tiempo de reacción promedio verdadero a un estímulo particular o es posible que un ingeniero de seguridad desee sólo un límite de confianza inferior para la vida útil promedio real de componentes de un tipo. Como el área acumulativa bajo la curva normal estándar a la izquierda de 1.645 es de .95,

$$P\left(\frac{\bar{X}-\mu}{S/\sqrt{n}} < 1.645\right) \approx .95$$

Si se manipula la desigualdad entre paréntesis para aislar μ en un lado y se reemplazan las variables aleatorias con valores calculados se obtiene la desigualdad $\mu > \bar{x} - 1.645s/\sqrt{n}$; la expresión a la derecha es el límite de confianza inferior deseado. Comenzando con $P(-1.645 < Z) \approx .95$ y manipulando la desigualdad se obtiene el límite de confianza superior. Un argumento similar da un límite unilateral asociado con cualquier otro nivel de confianza.

PROPOSICIÓN

Un **límite de confianza superior muestral grande para μ** es

$$\mu < \bar{x} + z_\alpha \cdot \frac{s}{\sqrt{n}}$$

y un **límite de confianza inferior muestral grande para μ** es

$$\mu > \bar{x} - z_\alpha \cdot \frac{s}{\sqrt{n}}$$

Se obtiene **un límite de confianza unilateral para p** reemplazando $z_{\alpha/2}$ en lugar de z_α y $\pm$ en lugar de $+$ o $-$ en la fórmula para el intervalo de confianza (7.10) para p. En todos los casos, el nivel de confianza es aproximadamente de $100(1-\alpha)\%$.

Ejemplo 7.10 La prueba de esfuerzo cortante es el procedimiento más aceptado para evaluar la calidad de una unión entre un material de reparación y su sustrato de concreto. El artículo "Testing the Bond Between Repair Materials and Concrete Substrate" (*ACI Materials J.*, 1996: 553–558) reportó que en una investigación particular, una muestra de 48 observaciones de resistencia al esfuerzo cortante dio una resistencia media muestral de 17.17 N/mm^2 y una desviación estándar muestral de 3.28 N/mm^2. Un límite de confianza inferior para la resistencia al esfuerzo cortante promedio verdadera μ con nivel de confianza de 95% es

$$17.17 - (1.645)\frac{(3.28)}{\sqrt{48}} = 17.17 - .78 = 16.39$$

Esto es, con un nivel de confianza de 95%, se puede decir que $\mu > 16.39$. ■

EJERCICIOS Sección 7.2 (12–27)

12. Una muestra aleatoria de 110 relámpagos en cierta región dieron por resultado una duración de eco de radar promedio muestral de .81 segundos y una desviación estándar muestral de .34 segundos ("Lightning Strikes to an Airplane in a Thunderstorm", *J. of Aircraft*, 1984: 607–611). Calcule un intervalo de confianza de 99% (bilateral) para la duración de eco promedio verdadera μ e interprete el intervalo resultante.

13. El artículo "Gas Cooking, Kitchen Ventilation, and Exposure to Combustion Products" (*Indoor Air*, 2006: 65–73) reportó que para una muestra de 50 cocinas con estufas de gas monitoreadas durante una semana, el nivel de CO_2 medio muestral (ppm) fue de 654.16 y la desviación estándar muestral fue de 164.43.

a. Calcule e interprete un intervalo de confianza de 95% (bilateral) para un nivel de CO_2 promedio verdadero en la población de todas las casas de la cual se seleccionó la muestra.

b. Suponga que el investigador había hecho una suposición preliminar de 175 para el valor de s antes de recopilar los datos. ¿Qué tamaño de muestra sería necesario para obtener un ancho de intervalo de 50 ppm para un nivel de confianza de 95%?

14. El artículo "Evaluating Tunnel Kiln Performance" (*Amer. Ceramic Soc. Bull.*, agosto de 1997: 59–63) reportó la siguiente información resumida sobre resistencias a la fractura (MPa) de $n = 169$ barras de cerámica horneadas en un horno particular: $\bar{x} = 89.10$, $s = 3.73$.

a. Calcule un intervalo de confianza (bilateral) para la resistencia a la fractura promedio verdadera utilizando un nivel de confianza de 95%. ¿Se podría decir que la resistencia a la fractura promedio verdadera fue estimada con precisión?

b. Suponga que los investigadores creyeron *a priori* que la desviación estándar de la población era aproximadamente de 4 MPa. Basado en esta suposición, ¿qué tan grande tendría que ser una muestra para estimar μ dentro de .5 MPa con 95% de confianza?

15. Determine el nivel de confianza de cada uno de los siguientes límites de confianza unilaterales muestrales grandes:

a. Límite superior: $\bar{x} + .84s/\sqrt{n}$

b. Límite inferior: $\bar{x} - 2.05s/\sqrt{n}$

c. Límite superior: $\bar{x} + .67s/\sqrt{n}$

16. El voltaje de ruptura de corriente alterna (AC) de un líquido aislante indica su rigidez dieléctrica. El artículo "Testing Practices for the AC Breakdown Voltage Testing of Insulation Liquids" (*IEEE Electrical Insulation Magazine*, 1995: 21−26) dio las observaciones muestrales adjuntas de voltaje de ruptura (kV) de un circuito particular, en ciertas condiciones.

62	50	53	57	41	53	55	61	59	64	50	53	64	62	50	68
54	55	57	50	55	50	56	55	46	55	53	54	52	47	47	55
57	48	63	57	57	55	53	59	53	52	50	55	60	50	56	58

a. Construya un diagrama de caja de los datos y comente sobre las características interesantes.

b. Calcule e interprete un intervalo de confianza del 95% para el promedio real del voltaje de ruptura μ. ¿Parece que μ ha sido estimada con precisión? Explique.

c. Supongamos que el investigador cree que prácticamente todos los valores de voltaje de ruptura están entre 40 y 70. ¿Qué tamaño de la muestra sería conveniente para que el intervalo de confianza del 95% tenga una anchura de 2 kV (de modo que μ se estime dentro de 1 kV con 95% de confianza)?

17. El ejercicio 1.13 dio una muestra de observaciones de resistencia última a la tensión (kg/pulg^2). Use los datos de salida estadísticos descriptivos adjuntos de Minitab para calcular un límite de confianza inferior de 99% para la resistencia a la tensión última promedio verdadera e interprete el resultado.

N	Media	Mediana	TrMedia	DesvEst	ECMedia
153	135.39	135.40	135.41	4.59	0.37

Mínimo	Máximo	Q1	Q3
122.20	147.70	132.95	138.25

18. El artículo "Ultimate Load Capacities of Expansion Anchor Bolts" (*J. of Energy Engr.*, 1993: 139–158) reportó los siguientes datos resumidos sobre resistencia al esfuerzo cortante (kg/pulg^2) para una muestra de pernos de anclaje de 3/8 pulg: $n = 78$, $\bar{x} = 4.25$, $s = 1.30$. Calcule un límite de confianza inferior utilizando un nivel de confianza de 90% para una resistencia al esfuerzo cortante promedio verdadera.

19. El artículo "Limited Yield Estimation for Visual Defect Sources" (*IEEE Trans. on Semiconductor Manuf.*, 1997: 17–23) reportó que, en un estudio de un proceso de inspección de obleas particular, 356 troqueles fueron examinados por una sonda de inspección y 201 de éstos pasaron la prueba. Suponiendo un proceso estable, calcule un intervalo de confianza (bilateral) de 95% para la proporción de todos los troqueles que pasan la prueba.

20. La Prensa Asociada (9 de octubre de 2002) reportó que en una encuesta de 4722 jóvenes estadounidenses de 6 a 19 años de edad, 15% sufría de problemas serios de sobrepeso (un índice de masa corporal de por lo menos 30; este índice mide el peso con respecto a la estatura). Calcule e interprete un intervalo de confianza utilizando un nivel de confianza de 99% para la proporción de todos los jóvenes estadounidenses con un problema serio de sobrepeso.

21. En una muestra de 1000 consumidores seleccionados al azar que tuvieron la oportunidad de enviar un formulario de solicitud de reembolso después de comprar un producto, 250 de estas personas dijeron que nunca lo hicieron ("Rebates: Get What You Deserve", *Consumer Reports*, mayo de 2009: 7). Las razones citadas para su comportamiento incluyen demasiados pasos en el proceso, cantidad demasiado pequeña, vencimiento del plazo, el temor de ser puesto en una lista de correo, la pérdida de su recepción y las dudas acerca de recibir el dinero. Calcule un límite de confianza superior al nivel de confianza del 95% para la verdadera proporción de estos consumidores que nunca solicitaron un reembolso. Con base en este límite, ¿hay pruebas convincentes de que la verdadera proporción de estos consumidores es menor que 1/3? Explique su razonamiento.

22. La tecnología subyacente de reemplazos de cadera ha cambiado ya que estas operaciones se han vuelto más populares (más de 250,000 en Estados Unidos en el año 2008). A partir del año 2003, las caderas de cerámica de alta duración se comercializaban. Desafortunadamente, para muchos pacientes la mayor durabilidad ha sido compensada por un aumento en la incidencia de chirridos. La edición del 11 de mayo de 2008 del *New York Times* informó que en un estudio de 143 individuos que recibieron las caderas de cerámica entre los años 2003 y 2005, 10 de las caderas desarrollaron chirridos.

a. Calcule un límite de confianza inferior en el nivel de confianza del 95% para la verdadera proporción de las caderas que desarrollaron chirridos.

b. Interprete el nivel de confianza del 95% utilizado en el inciso (a).

23. El Pew Forum on Religion and Public Life reportó el 9 de diciembre del año 2009 que en una encuesta de 2003 adultos estadounidenses, 25% dijo que creía en la astrología.

a. Calcule e interprete un intervalo de confianza al nivel de confianza del 99% para la proporción de todos los adultos estadounidenses que creen en la astrología.

b. ¿Qué tamaño de muestra se requiere para que el ancho de un intervalo de confianza de 99% tenga un máximo de .05, independientemente del valor de $\hat{p}$?

24. Una muestra de 56 muestras de algodón produjo un porcentaje de alargamiento promedio muestral de 8.17 y una desviación estándar de 1.42 ("An Apparent Relation Between the Spiral Angle ϕ, the Percent Elongation E_1, and the Dimensions of the Cotton Fiber", *Textile Research J.*, 1978: 407–410). Calcule un intervalo de confianza de 95% muestral grande para el porcentaje

de alargamiento promedio verdadero μ. ¿Qué suposiciones está haciendo sobre la distribución del porcentaje de alargamiento?

25. Una legisladora estatal desea encuestar a los residentes de su distrito para ver qué proporción del electorado está consciente de su posición sobre la utilización de fondos estatales para solventar abortos.

a. ¿Qué tamaño de muestra es necesario si el intervalo de confianza de 95% para p tiene que tener un ancho de cuando mucho .10 independientemente de p?

b. Si la legisladora está firmemente convencida de que por lo menos 2/3 del electorado conoce su posición, ¿qué tamaño de muestra recomendaría?

26. El superintendente de un gran distrito escolar, que una ocasión tomó un curso de probabilidad y estadística, cree que el número de maestros ausentes en cualquier día dado tiene una distribución de Poisson con parámetro μ. Use los datos adjuntos sobre ausencias durante 50 días para obtener un intervalo de confianza muestral grande para μ. [*Sugerencia*: la media y la varianza de una variable de Poisson son iguales a μ, por consiguiente

$$Z = \frac{\overline{X} - \mu}{\sqrt{\mu/n}}$$

tiene aproximadamente una distribución normal estándar. Ahora prosiga como en la deducción del intervalo para p haciendo una proposición de probabilidad (con probabilidad de $1 - \alpha$) y resolviendo las desigualdades resultantes para μ (véase el argumento exactamente después de (7.10).]

Número de ausencias	0	1	2	3	4	5	6	7	8	9	10
Frecuencia	1	4	8	10	8	7	5	3	2	1	1

27. Reconsidere el intervalo de confianza (7.10) para p y enfóquese en un nivel de confianza de 95%. Demuestre que los límites de confianza concuerdan bastante bien con los del intervalo tradicional (7.11) una vez que dos éxitos y dos fallas se anexaron a la muestra [es decir, (7.11) basado en $x + 2$ S en $n + 4$ ensayos]. [*Sugerencia*: $1.96 \approx 2$. *Nota*: Agresti y Coull demostraron que este ajuste del intervalo tradicional también tiene un nivel de confianza próximo al nivel nominal.]

7.3 Intervalos basados en una distribución de población normal

El intervalo de confianza para μ presentado en la sección 7.2 es válido siempre que n es grande. El intervalo resultante puede ser utilizado cualquiera que sea la naturaleza de la distribución de la población. El teorema del límite central no puede ser invocado, sin embargo, cuando n es pequeña. En este caso, una forma de proceder es hacer una suposición específica sobre la forma de la distribución de la población y luego deducir un intervalo de confianza adecuado a esa suposición. Por ejemplo, se podría desarrollar un intervalo de confianza para μ, cuando una distribución gamma describe la población, otro para el caso de una población de Weibull, y así sucesivamente. Estadísticos en realidad han realizado este programa para varias familias distribucionales diferentes. Como la distribución normal es más frecuentemente apropiada como modelo de población que cualquier otro tipo de distribución, la atención aquí se concentrará en un intervalo de confianza para esta situación.

SUPOSICIÓN

La población de interés es normal, de modo que $X_1, \ldots, X_n$ constituyen una muestra aleatoria tomada de una distribución normal con μ y σ desconocidas.

El resultado clave que subyace en el intervalo de la sección 7.2 fue que con n grande, la variable aleatoria $Z = (\overline{X} - \mu)/(S/\sqrt{n})$ tiene aproximadamente una distribución normal estándar. Cuando n es pequeño, no es probable que S se aproxime a σ, de modo que la variabilidad de la distribución de Z surge de la aleatoriedad tanto en el numerador como en el denominador. Esto implica que la distribución de probabilidad de $(\overline{X} - \mu)/(S/\sqrt{n})$ se dispersará más que la distribución normal estándar. El resultado en el cual están basadas las inferencias introduce una nueva familia de distribuciones de probabilidad llamada *distribuciones t*.

TEOREMA

Cuando $\overline{X}$ es la media de una muestra aleatoria de tamaño n tomada de una distribución normal con media μ, la variable aleatoria

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \tag{7.13}$$

tiene una distribución de probabilidad llamada distribución t con $n - 1$ grados de libertad (gl).

Propiedades de distribuciones *t*

Antes de aplicar este teorema, se impone una discusión de las propiedades de distribuciones t. Aunque la variable de interés sigue siendo $(\overline{X} - \mu)/(S/\sqrt{n})$, ahora se denota por T para recalcar que no tiene una distribución normal estándar cuando n es pequeña. Recuérdese que una distribución normal está regida por dos parámetros, cada elección diferente de μ en combinación con σ resulta en una distribución normal particular. Cualquier distribución t particular resulta de especificar el valor de sólo un parámetro, llamado **número de grados de libertad**, abreviado como gl. Este parámetro se denota con la letra griega ν. Posibles valores de ν son los enteros positivos 1, 2, 3, . . . Así que hay una distribución t con un gl, otra con 2 gl, otra con 3 gl, y así sucesivamente.

Para cualquier valor fijo del parámetro ν, la función de densidad que especifica la curva t asociada es incluso más complicada que la función de densidad normal. Afortunadamente, sólo hay que ocuparse de algunas de las más importantes características de estas curvas.

Propiedades de distribuciones *t*

Sea t_ν, que denota la distribución t con ν gl.

1. Cada curva t_ν tiene forma de campana y su centro en 0.
2. Cada curva t_ν está más esparcida que la curva (z) normal estándar.
3. Conforme ν se incrementa, la dispersión de la curva t_ν correspondiente disminuye.
4. A medida que $\nu \to \infty$, la secuencia de curvas t_ν tiende a la curva normal estándar (así que la curva z a menudo se llama curva t con grados de libertad $= \infty$).

La figura 7.7 ilustra varias de estas propiedades para valores seleccionados de ν.

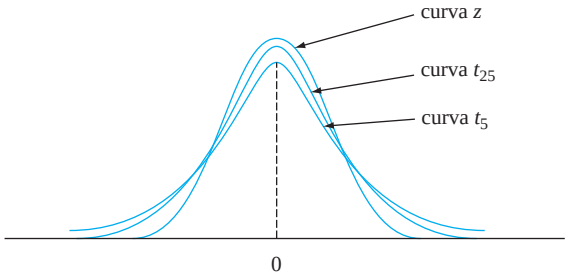


Figura 7.7 Curvas t_ν y z

El número de grados de libertad para T en (7.13) es $n - 1$ porque, aunque S está basada en las n desviaciones $X_1 - \overline{X}, \ldots, X_n - \overline{X}, \Sigma(X_i - \overline{X}) = 0$ implica que sólo

$n - 1$ de éstas están "libremente determinadas". El número de grados de libertad para una variable t es el número de desviaciones libremente determinadas en las cuales está basada la desviación estándar estimada en el denominador de T.

El uso de la distribución t al hacer inferencias requiere notación para capturar áreas de cola de la curva t análogas a z_α de la curva z. Se podría pensar que t_α haría el truco. Sin embargo, el valor deseado depende no sólo del área de la cola capturada, sino también de gl.

NOTACIÓN

Sea $t_{\alpha,\nu}$ = el número sobre el eje de medición con el cual el área bajo la curva t con ν grados de libertad a la derecha de $t_{\alpha,\nu}$ es α; $t_{\alpha,\nu}$ se llama **valor crítico *t***.

Por ejemplo, $t_{.05,6}$ es el valor crítico t que captura un área de cola superior de .05 bajo la curva t con 6 gl. La notación general se ilustra en la figura 7.8. Debido a que las curvas t son simétricas alrededor de cero, $-t_{\alpha,\nu}$ captura el área α de la cola inferior. La tabla A.5 del apéndice da $t_{\alpha,\nu}$ para valores seleccionados de α y ν. Esta tabla también aparece al final del libro. Las columnas de la tabla corresponden a diferentes valores de α. Para obtener $t_{.05,15}$, vaya a la columna $\alpha = .05$, mire hacia abajo al renglón $\nu = 15$, y lea $t_{.05,15} = 1.753$. Del mismo modo, $t_{.05,22} = 1.717$ (columna .05, renglón $\nu = 22$) y $t_{.01,22} = 2.508$.

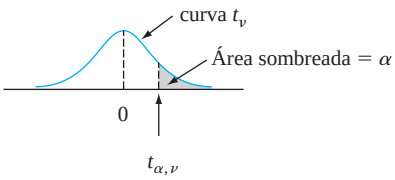


Figura 7.8 Ilustración de un valor crítico t

Los valores de $t_{\alpha,\nu}$ exhiben un comportamiento regular al recorrer una fila o al descender por una columna. Con ν fijo, $t_{\alpha,\nu}$ se incrementa a medida que α disminuye, puesto que hay que moverse más a la derecha de cero para capturar el área α en la cola. Con α fija, a medida que ν se incrementa (es decir, cuando se recorre hacia abajo cualquier columna particular de la tabla t) el valor de $t_{\alpha,\nu}$ disminuye. Esto es porque un valor más grande de ν implica una distribución t con dispersión más pequeña, de modo que no es necesario ir más lejos de cero para capturar el área de cola α. Además, $t_{\alpha,\nu}$ disminuye más lentamente a medida que ν se incrementa. Por consiguiente, los valores que aparecen en la tabla se muestran en incrementos de 2 entre 30 grados de libertad y 40 grados de libertad y luego saltan a $\nu = 50$, 60, 120 y por último ∞. Como t_∞ es la curva normal estándar, los valores z_α conocidos aparecen en la última fila de la tabla. La regla empírica sugería con anterioridad que el uso del intervalo de confianza muestral grande (si $n > 40$) proviene de la igualdad aproximada de las distribuciones normales estándar y t para $\nu \geq 40$.

Intervalo de confianza *t* para una muestra

La variable estandarizada T tiene una distribución t con $n - 1$ grados de libertad y el área bajo la curva de densidad t correspondiente entre $-t_{\alpha/2,n-1}$ y $t_{\alpha/2,n-1}$ es $1 - \alpha$ (el área $\alpha/2$ queda en cada cola), por consiguiente

$$P(-t_{\alpha/2,\,n-1} < T < t_{\alpha/2,\,n-1}) = 1 - \alpha \tag{7.14}$$

La expresión (7.14) difiere de las expresiones que aparecen en secciones previas en que T y $t_{\alpha/2,n-1}$ se utilizan en lugar de Z y $z_{\alpha/2}$, aunque pueden ser manipuladas de la misma manera para obtener un intervalo de confianza para μ.

PROPOSICIÓN

Sean $\bar{x}$ y s la media y la desviación estándar muestrales calculadas con los resultados de una muestra aleatoria tomada de una población normal con media μ. Entonces un **intervalo de confianza de 100(1 − α)% para μ** es

$$\left(\bar{x} - t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}},\ \bar{x} + t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}}\right) \tag{7.15}$$

o, más compactamente, $\bar{x} \pm t_{\alpha/2,n-1} \cdot s/\sqrt{n}$.

Un **límite de confianza superior para μ** es

$$\bar{x} + t_{\alpha,n-1} \cdot \frac{s}{\sqrt{n}}$$

y reemplazando + por − en la última expresión se obtiene un **límite de confianza inferior para μ,** ambos con nivel de confianza de 100(1 − α)%.

Ejemplo 7.11 A pesar de que los mercados tradicionales de madera de ocozol han disminuido, las maderas sólidas de gran sección utilizadas tradicionalmente para la construcción de puentes y duelas se han vuelto cada vez más escasas. El artículo "Development of Novel Industrial Laminated Planks from Sweetgum Lumber" (*J. of Bridge Engr.*, 2008: 64–66) describe la fabricación y el ensayo de vigas compuestas, concebida para agregar valor a la madera de ocozol de bajo grado. Aquí hay datos sobre el módulo de ruptura (psi; el artículo contenía datos resumidos, expresados en MPa):

6807.99	7637.06	6663.28	6165.03	6991.41	6992.23
6981.46	7569.75	7437.88	6872.39	7663.18	6032.28
6906.04	6617.17	6984.12	7093.71	7659.50	7378.61
7295.54	6702.76	7440.17	8053.26	8284.75	7347.95
7422.69	7886.87	6316.67	7713.65	7503.33	7674.99

La figura 7.9 muestra un diagrama de probabilidad normal obtenido con el software R. La derechura del patrón en el diagrama apoya fuertemente la suposición de que la distribución de la población del módulo de ruptura es por lo menos aproximadamente normal.

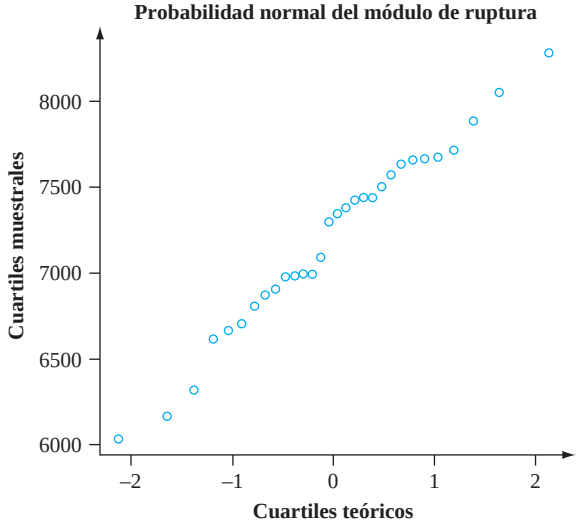


Figura 7.9 Diagrama de probabilidad normal de los datos del módulo de ruptura

La media muestral y la desviación estándar de la muestra son 7203.191 y 543.5400, respectivamente (para abatir cualquier realización de cálculos a mano, la carga computacional se alivia un poco al restar 6000 de cada valor de x para obtener $y_i = x_i - 6000$, entonces $\sum y_i = 36{,}095.72$ y $\sum y_i^2 = 51{,}997{,}668.77$, de la cual $\bar{y} = 1203.191$ y $s_y = s_x$ tal como se indica).

Ahora se calcula un intervalo de confianza para el promedio real del módulo de ruptura con un nivel de confianza de 95%. El intervalo de confianza se basa en $n - 1 = 29$ grados de libertad, por lo que el valor de t crítico necesario es $t_{.025,29} = 2.045$. La estimación por intervalo es ahora

$$\bar{x} \pm t_{.025,29} \cdot \frac{s}{\sqrt{n}} = 7203.191 \pm (2.045) \cdot \frac{543.5400}{\sqrt{30}}$$
$$= 7203.191 \pm 202.938 = (7000.253, 7406.129)$$

Se estima que $7000.253 < \mu < 7406.129$ con un 95% de confianza. Si se utiliza la misma fórmula en muestra tras muestra, en el largo plazo el 95% de los intervalos calculados contendrán a μ. Dado que el valor de μ no está disponible, no sabemos si el intervalo calculado es uno de los "buenos" del 95% o el "malo" del 5%. Incluso con el tamaño de la muestra moderadamente grande, el intervalo es bastante amplio. Esto es una consecuencia de la cantidad sustancial de variabilidad de la muestra en los valores del módulo de ruptura.

Un límite de confianza inferior al 95% resultaría de conservar únicamente el límite de confianza inferior (el que tiene el signo $(-)$) y reemplazar 2.045 con $t_{.05,29} = 1.699$. ■

Por desgracia, no es fácil seleccionar n para controlar el ancho del intervalo t. Esto es porque el ancho implica la s desconocida (antes de recopilar los datos) y porque n ingresa no sólo a través de $1/\sqrt{n}$ sino también a través de $t_{\alpha/2,n-1}$. Por consiguiente, se puede obtener una n apropiada sólo mediante ensayo y error.

En el capítulo 15 se discutirá un intervalo de confianza de muestra pequeña para μ que es válido siempre que la distribución de la población sea simétrica, una suposición más débil que la de normalidad. No obstante, cuando la distribución de la población es normal, el intervalo t tiende a acortarse más de lo que lo haría *cualquier* otro intervalo con el mismo nivel de confianza.

Un intervalo de predicción para un solo valor futuro

En muchas aplicaciones, el objetivo es *predecir* un solo valor de una variable que tiene que ser observada en un tiempo futuro, en lugar de *estimar* el valor medio de dicha variable.

Ejemplo 7.12 Considere la siguiente muestra de contenido de grasa (en porcentaje) de $n = 10$ perros calientes seleccionados al azar ("Sensory and Mechanical Assessment of the Quality of Frankfurters", *J. of Texture Studies*, 1990: 395–409):

25.2 21.3 22.8 17.0 29.8 21.0 25.5 16.0 20.9 19.5

Suponiendo que estas observaciones se seleccionaron de una distribución de población normal, un intervalo de confianza de 95% para (estimación del intervalo de) el contenido de grasa medio de la población es

$$\bar{x} \pm t_{.025,9} \cdot \frac{s}{\sqrt{n}} = 21.90 \pm 2.262 \cdot \frac{4.134}{\sqrt{10}} = 21.90 \pm 2.96$$
$$= (18.94, 24.86)$$

Suponga, sin embargo, que se va a comer un solo perro caliente de este tipo y desea *predecir* el contenido de grasa resultante. Una predicción *puntual*, análoga a una estimación *puntual*, es simplemente $\bar{x} = 21.90$. Esta predicción desafortunadamente no da información sobre confiabilidad o precisión. ■

El escenario general es como sigue: se dispone de una muestra aleatoria X_1, X_2, . . . , X_n tomada de una distribución de población normal y se desea predecir el valor de X_{n+1}, una sola observación futura (por ejemplo, la vida útil de un foco sencillo que se compra o la eficiencia de combustible de un automóvil simple que es rentado). Un predictor puntual es $\overline{X}$ y el error de predicción resultante es $\overline{X} - X_{n+1}$. El valor esperado del error de predicción es

$$E(\overline{X} - \overline{X}_{n+1}) = E(\overline{X}) - E(X_{n+1}) = \mu - \mu = 0$$

Como X_{n+1}, es independiente de $X_1, \ldots, X_n$, es independiente de $\overline{X}$, así que la varianza del error de predicción es

$$V(\overline{X} - X_{n+1}) = V(\overline{X}) + V(X_{n+1}) = \frac{\sigma^2}{n} + \sigma^2 = \sigma^2\left(1 + \frac{1}{n}\right)$$

El error de predicción es una combinación lineal de variables aleatorias independientes normalmente distribuidas, así que también está normalmente distribuido. Por consiguiente

$$Z = \frac{(\overline{X} - X_{n+1}) - 0}{\sqrt{\sigma^2\left(1 + \frac{1}{n}\right)}} = \frac{\overline{X} - X_{n+1}}{\sqrt{\sigma^2\left(1 + \frac{1}{n}\right)}}$$

tiene una distribución normal estándar. Se puede demostrar que si se reemplaza σ con la desviación estándar muestral S (de $X_1, \ldots, X_n$) se obtiene

$$T = \frac{\overline{X} - X_{n+1}}{S\sqrt{1 + \frac{1}{n}}} \sim t \text{ distribución } t \text{ con } n - 1 \text{ grados de libertad}$$

Si se manipula esta variable T como se manipuló $T = (\overline{X} - \mu)/(S/\sqrt{n})$ en el desarrollo de un intervalo de confianza se obtiene el siguiente resultado.

PROPOSICIÓN

Un **intervalo de predicción** (IP) para una sola observación que tiene que ser seleccionado de una distribución de población normal es

$$\bar{x} \pm t_{\alpha/2,n-1} \cdot s\sqrt{1 + \frac{1}{n}} \qquad (7.16)$$

El *nivel de predicción* es $100(1 - \alpha)\%$. Una predicción del límite inferior resulta de la sustitución de $t_{\alpha/2}$ por t_α y desechar la parte + de (7.16), una modificación similar da una predicción del límite superior.

La interpretación de un nivel de predicción de 95% es similar a la de un nivel de confianza de 95%; si se calcula el intervalo (7.16) para muestra tras muestra, a la larga el 95% de estos intervalos incluirán los valores futuros correspondientes de X.

Ejemplo 7.13 (Continuación del ejemplo 7.12)

Con $n = 10$, $\bar{x} = 21.90$, $s = 4.134$ y $t_{.025,9} = 2.262$, un intervalo de predicción de 95% para el contenido de grasa de un solo perro caliente es

$$21.90 \pm (2.262)(4.134)\sqrt{1 + \frac{1}{10}} = 21.90 \pm 9.81$$
$$= (12.09, 31.71)$$

El intervalo es bastante ancho, lo que indica una incertidumbre sustancial en cuanto al contenido de grasa. Obsérvese que el ancho del intervalo de predicción es más de tres veces el del intervalo de confianza. ■

El error de predicción es $\bar{X} - X_{n+1}$, la diferencia entre dos variables aleatorias, en tanto que el error de estimación es $\bar{X} - \mu$, la diferencia entre una variable aleatoria y un valor fijo (aunque desconocido). El intervalo de predicción es más ancho que el intervalo de confianza porque hay más variabilidad en el error de predicción (debido a X_{n+1}) que en el error de estimación. De hecho, a medida que n se hace arbitrariamente grande, el intervalo de confianza se contrae a un solo valor μ y el intervalo de predicción tiende a $\mu \pm z_{\alpha/2} \cdot \sigma$. Existe incertidumbre con respecto a un solo valor X incluso cuando no hay necesidad de estimarlo.

Intervalos de tolerancia

Considérese una población de automóviles de cierto tipo y supóngase que en condiciones específicas, la eficiencia de combustible (mpg) tiene una distribución normal con $\mu = 30$ y $\sigma = 2$. Entonces como el intervalo de -1.645 a 1.645 captura 90% del área bajo la curva z, 90% de todos estos automóviles tendrán valores de eficiencia de combustible entre $\mu - 1.645\sigma = 26.71$ y $\mu + 1.645\sigma = 33.29$. Pero, ¿qué sucederá si los valores de μ y σ no son conocidos? Se puede tomar una muestra de tamaño n, determinar las eficiencias de combustible, $\bar{x}$ y s, y formar el intervalo cuyo límite inferior es $\bar{x} - 1.645s$ y cuyo límite superior es $\bar{x} + 1.645s$. Sin embargo, debido a la variabilidad de muestreo en las estimaciones de μ y σ, existe una buena probabilidad de que el intervalo resultante incluya menos de 90% de los valores de la población. Intuitivamente, para tener *a priori* una probabilidad de que 95% del intervalo resultante incluya por lo menos 90% de los valores de la población, cuando $\bar{x}$ y s se utilizan en lugar de μ y σ, también se deberá reemplazar 1.645 con un número más grande. Por ejemplo, cuando $n = 20$, el valor 2.310 es tal que se puede estar 95% confiado en que el intervalo $\bar{x} \pm 2.310s$ incluirá por lo menos 90% de los valores de eficiencia de combustible en la población.

Sea k un número entre 0 y 100. Un **intervalo de tolerancia** para capturar por lo menos el k% de los valores en una distribución de población normal con nivel de confianza de 95% tiene la forma

$$\bar{x} \pm (\text{valor crítico de tolerancia}) \cdot s$$

En la tabla A.6 del apéndice aparecen valores críticos de tolerancia para $k = 90$, 95 y 99 en combinación con varios tamaños de muestra. Esta tabla también incluye valores críticos para un nivel de confianza de 99% (estos valores son más grandes que los valores correspondientes al 95%). Si se reemplaza $\pm$ con $+$ se obtiene un límite de tolerancia superior, y si se utiliza $-$ en lugar de $\pm$ se obtiene un límite de tolerancia inferior. En la tabla A.6 también aparecen valores críticos para obtener estos límites unilaterales.

Ejemplo 7.14

Como parte de un proyecto más amplio para estudiar el comportamiento de los paneles de corteza comprimida, un componente estructural que se utiliza ampliamente en América del Norte, el artículo "Time-Dependent Bending Properties of Lumber" (*J. of Testing and Eval.*, 1996: 187–193) informó sobre varias propiedades mecánicas de muestras de madera de pino escocés. Considere las siguientes observaciones sobre el módulo de elasticidad (MPa) obtenido un minuto después de la carga en una cierta configuración:

10,490	16,620	17,300	15,480	12,970	17,260	13,400	13,900
13,630	13,260	14,370	11,700	15,470	17,840	14,070	14,760

Hay un patrón lineal pronunciado en el gráfico de probabilidad normal de los datos. Un resumen de las cantidades importantes es $n = 16$, $\bar{x} = 14{,}532.5$, $s = 2055.67$. Para un nivel de confianza del 95%, un intervalo de tolerancia bilateral para la captura de al menos

el 95% de los valores de los módulos elasticidad de las muestras de madera en la población de la muestra, utiliza el valor de tolerancia crítica de 2.903. El intervalo resultante es

$$14{,}532.5 \pm (2.903)(2055.67) = 14{,}532.5 \pm 5967.6 = (8{,}564.9,\ 20{,}500.1)$$

Se puede estar totalmente confiado de que por lo menos 95% de todos los especímenes de madera tienen valores de módulo de elasticidad de entre 8564.9 y 20,500.1.

El intervalo de confianza de 95% para μ fue (13,437.3, 15,627.7) y el intervalo de predicción de 95% para el módulo de elasticidad de un solo espécimen de madera es (10,017.0, 19,048.0). Tanto el intervalo de predicción como el intervalo de tolerancia son sustancialmente más anchos que el intervalo de confianza. ■

Intervalos basados en distribuciones de población no normales

El intervalo de confianza *t* para una muestra de μ es robusto en cuanto a alejamientos pequeños o incluso moderados de la normalidad a menos que *n* sea bastante pequeño. Con esto se quiere decir que si se utiliza un valor crítico para confianza de 95%, por ejemplo, al calcular el intervalo, el nivel de confianza real se aproximará de manera razonable al nivel nominal de 95%. Si, sin embargo, *n* es pequeño y la distribución de la población es altamente no normal, entonces el nivel de confianza real puede ser diferente en forma considerable del que se utiliza cuando se obtiene un valor crítico particular de la tabla *t*. Ciertamente, ¡sería penoso creer que el nivel de confianza es de más o menos 95% cuando en realidad es como de 88%! Se ha visto que la técnica bootstrap, introducida en la sección 7.1 es bastante exitosa al estimar parámetros en una amplia variedad de situaciones no normales.

En contraste con el intervalo de confianza, la validez de los intervalos de predicción y tolerancia descritos en esta sección está estrechamente vinculada a la suposición de normalidad. Estos últimos intervalos no deberán ser utilizados sin evidencia apremiante de normalidad. La excelente referencia *Statistical Intervals*, citada en la bibliografía al final de este capítulo, discute procedimientos alternativos de esta clase en varias otras situaciones.

EJERCICIOS Sección 7.3 (28–41)

28. Determine los valores de las siguientes cantidades:
a. $t_{.1,15}$ b. $t_{.05,15}$ c. $t_{.05,25}$ d. $t_{.05,40}$ e. $t_{.005,40}$

29. Determine el valor o valores crítico(os) *t* que capturará el área deseada de la curva *t* en cada uno de los siguientes casos:
a. Área central = .95, gl = 10
b. Área central = .95, gl = 20
c. Área central = .99, gl = 20
d. Área central = .99, gl = 50
e. Área de cola superior = .01, gl = 25
f. Área de cola inferior = .025, gl = 5

30. Determine el valor crítico *t* de un intervalo de confianza bilateral en cada una de las siguientes situaciones:
a. Nivel de confianza = 95%, gl = 10
b. Nivel de confianza = 95%, gl = 15
c. Nivel de confianza = 99%, gl = 15
d. Nivel de confianza = 99%, $n = 5$
e. Nivel de confianza = 98%, gl = 24
f. Nivel de confianza = 99%, $n = 38$

31. Determine el valor crítico *t* para un límite de confianza inferior o superior en cada una de las situaciones descritas en el ejercicio 30.

32. De acuerdo con el artículo “Fatigue Testing of Condoms” (*Polymer Testing*, 2009: 567−571), “las pruebas que se utilizan actualmente para los condones son sustitutos de los desafíos que enfrentan en uso”, incluyendo una prueba de hoyos, una prueba de inflación, una prueba de sello del paquete y las pruebas de las dimensiones y la calidad del lubricante (¡todo el territorio fértil para el uso de la metodología estadística!). Los investigadores desarrollaron una nueva prueba que agrega tensión cíclica a un nivel muy por debajo de la rotura y determina el número de ciclos hasta llegar a la ruptura. Una muestra de 20 condones de un tipo particular, resultó en una media muestral de 1584 y una desviación estándar muestral de 607. Calcule e interprete un intervalo de confianza al nivel de confianza del 99% para el verdadero número promedio de ciclos de ruptura. [*Nota*: el artículo presenta los resultados de las pruebas de hipótesis basadas en la distribución *t*, la validez de éstas depende de suponer la distribución normal de la población.]

33. El artículo “Measuring and Understanding the Aging of Kraft Insulating Paper in Power Transformers” (*IEEE Electrical Insul. Mag.*, 1996: 28–34) contiene las siguientes observaciones del grado de polimerización de especímenes de papel para

los cuales la concentración de tiempos de viscosidad cayeron en un rango medio:

418 421 421 422 425 427 431
434 437 439 446 447 448 453
454 463 465

a. Construya una gráfica de caja de los datos y comente sobre cualquier característica interesante.
b. ¿Es plausible que las observaciones muestrales dadas fueran seleccionadas de una distribución normal?
c. Calcule un intervalo de confianza de 95% bilateral para un grado de polimerización promedio verdadero (como lo hicieron los autores del artículo). ¿Sugiere este intervalo que 440 es un valor factible del grado de polimerización promedio verdadero? ¿Qué hay en cuanto a 450?

34. Una muestra de 14 especímenes de junta de un tipo particular produjo un esfuerzo límite proporcional medio muestral de 8.48 MPa y una desviación estándar muestral de .79 MPa ("Characterization of Bearing Strength Factors in Pegged Timber Connections", *J. of Structural Engr*., 1997: 326–332).
a. Calcule e interprete un límite de confianza inferior de 95% para el esfuerzo límite proporcional promedio verdadero de todas las juntas. ¿Qué suposiciones, si hay alguna, hizo sobre la distribución del esfuerzo límite proporcional?
b. Calcule e interprete un límite de predicción inferior de 95% para el esfuerzo límite proporcional de una sola unión de este tipo.

35. Para corregir deformidades nasales congénitas se utiliza rinoplastia de aumento mediante implante de silicón. El éxito del procedimiento depende de varias propiedades biomecánicas del periostio y fascia nasales humanas. El artículo "Biomechanics in Augmentation Rhinoplasty" (*J. of Med. Engr. and Tech.*, 2005: 14-17) reportó que para una muestra de 15 adultos (recién fallecidos), la deformación de falla media (en porcentaje) fue de 25.0 y la desviación estándar fue de 3.5.
a. Suponiendo una distribución normal de la deformación de falla, estime la deformación promedio verdadera en una forma que transmita información acerca de precisión y confiabilidad.
b. Pronostique la deformación para un solo adulto en una forma que transmita información sobre precisión y confiabilidad. ¿Cómo se compara la predicción con la estimación calculada en el inciso (a)?

36. Las $n = 26$ observaciones de tiempo de escape dadas en el ejercicio 36 del capítulo 1 dan una media y desviación estándar muestrales de 370.69 y 24.36, respectivamente.
a. Calcule un límite de confianza superior para el tiempo de escape medio de la población utilizando un nivel de confianza de 95%.
b. Calcule un límite de predicción superior para el tiempo de escape de un solo trabajador adicional utilizando un nivel de predicción de 95%. ¿Cómo se compara este límite con el límite de confianza del inciso (a)?
c. Suponga que se escogerán dos trabajadores más para participar en el ejercicio de escape simulado. Denote sus tiempos de escape por X_{27} y X_{28} y sea $\overline{X}_{\text{nuevo}}$ el promedio de estos dos valores. Modifique la fórmula para un intervalo de predicción con un solo valor de x para obtener un intervalo de predicción para $\overline{X}_{\text{nuevo}}$ y calcule un intervalo bilateral de 95% basado en los datos de escape dados.

37. Un estudio de la capacidad de individuos de caminar en línea recta ("Can We Really Walk Straight?" *Amer. J. of Physical Anthro*., 1992: 19-27) reportó los datos adjuntos sobre cadencia (pasos por segundo) con una muestra de $n = 20$ hombres saludables seleccionados al azar.

.95 .85 .92 .95 .93 .86 1.00 .92 .85 .81
.78 .93 .93 1.05 .93 1.06 1.06 .96 .81 .96

Un diagrama de probabilidad normal apoya de manera sustancial la suposición de que la distribución de la población de cadencia es aproximadamente normal. A continuación se da un resumen descriptivo de los datos obtenidos con Minitab:

```
Cadencia N   Media  Mediana  TrMedia  DesvEst  ECMedia
variable 20 0.9255  0.9300   0.9261   0.0809   0.0181

Cadencia      Mín      Máx       Q1       Q3
variable    0.7800   1.0600   0.8525   0.9600
```

a. Calcule e interprete un intervalo de confianza de 95% para la cadencia media de la población.
b. Calcule e interprete un intervalo de predicción de 95% para la cadencia de un solo individuo seleccionado al azar de esta población.
c. Calcule un intervalo que incluya por lo menos 99% de las cadencias en la distribución de la población utilizando un nivel de confianza de 95%.

38. Se seleccionó una muestra de 25 piezas de laminado utilizado en la fabricación de tarjetas de circuito y se determinó la cantidad de pandeo (pulg) en condiciones particulares para cada pieza y el resultado fue un pandeo medio muestral de .0635 y una desviación estándar muestral de .0065.
a. Calcule una predicción de la cantidad de pandeo de una sola pieza de laminado en una manera que proporcione información sobre precisión y confiabilidad.
b. Calcule un intervalo con el cual pueda tener un alto grado de confianza de que por lo menos 95% de todas las piezas de laminado produzcan cantidades de pandeo que estén entre los dos límites del intervalo.

39. El ejercicio 72 del capítulo 1 dio las siguientes observaciones en una medición de afinidad de receptor (volumen de distribución ajustado) con una muestra de 13 individuos sanos: 23, 39, 40, 41, 43, 47, 51, 58, 63, 66, 67, 69, 72.
a. ¿Es plausible que la distribución de la población de la cual se seleccionó esta muestra sea normal?
b. Calcule un intervalo con el cual pueda estar 95% confiado de que por lo menos 95% de todos los individuos saludables en la población tienen volúmenes de distribución ajustados que quedan entre los límites del intervalo.
c. Pronostique el volumen de distribución ajustado de un solo individuo saludable calculando un intervalo de predicción de 95%. ¿Cómo se compara el ancho de este intervalo con el ancho del intervalo calculado en el inciso (b)?

40. El ejercicio 13 del capítulo 1 presentó una muestra de $n = 153$ observaciones de resistencia última a la tensión y el ejercicio 17 de la sección previa dio cantidades resumidas y solicitó un intervalo de confianza muestral grande. Como el tamaño de la mues-

tra es grande, no se requieren suposiciones sobre la distribución de la población en cuanto a la validez del intervalo de confianza.

a. ¿Se requiere alguna suposición sobre la distribución de la resistencia a la tensión antes de calcular un límite de predicción inferior para la resistencia a la tensión del nuevo espécimen seleccionado por medio del método descrito en esta sección? Explique.

b. Use un paquete de software estadístico para investigar la factibilidad de una distribución de población normal.

c. Calcule un límite de predicción inferior con un nivel de predicción de 95% para la resistencia última a la tensión del siguiente espécimen seleccionado.

41. Una tabulación más extensa de valores críticos t que la que aparece en este libro muestra que para la distribución t con 20 grados de libertad, las áreas a la derecha de los valores .687, .860 y 1.064 son .25, .20 y .15, respectivamente. ¿Cuál es el nivel de confianza de cada uno de los siguientes tres intervalos de confianza para la media μ de una distribución de población normal? ¿Cuál de los tres intervalos recomendaría utilizar y por qué?

a. $(\bar{x} - .687s/\sqrt{21}, \bar{x} + 1.725s/\sqrt{21})$

b. $(\bar{x} - .860s/\sqrt{21}, \bar{x} + 1.325s/\sqrt{21})$

c. $(\bar{x} - 1.064s/\sqrt{21}, \bar{x} + 1.064s/\sqrt{21})$

7.4 Intervalos de confianza para la varianza y desviación estándar de una población normal

Aun cuando las inferencias en lo que se refiere a la varianza σ^2 o a la desviación estándar de una población en general son de menor interés que aquellas con respecto a una media o proporción, hay ocasiones en que se requieren tales procedimientos. En el caso de una distribución de población normal, las inferencias están basadas en el siguiente resultado por lo que se refiere a la varianza muestral S^2.

TEOREMA

Sean $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución normal con parámetros μ y σ^2. Entonces la variable aleatoria

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum(X_i - \bar{X})^2}{\sigma^2}$$

tiene una distribución de probabilidad ji cuadrada (χ^2) con $n - 1$ grados de libertad.

Como se discutió en las secciones 4.4 y 7.1, la distribución ji cuadrada es una distribución de probabilidad continua con un solo parámetro ν, llamado número de grados de libertad, con posibles valores de 1, 2, 3, . . . Las gráficas de varias funciones de densidad de probabilidad χ^2 se ilustran en la figura 7.10. Cada función de densidad de probabilidad $f(x; \nu)$ es positiva sólo para $x > 0$, y cada una tiene asimetría positiva (una larga cola superior), aunque la distribución se mueve hacia la derecha y se vuelve más simétrica a medida que se incrementa ν. Para especificar procedimientos inferenciales que utilizan la distribución ji cuadrada, se requiere una notación análoga a aquella para un valor t crítico $t_{\alpha,\nu}$.

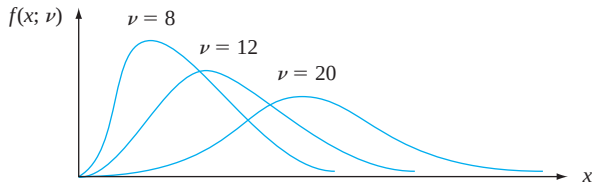


Figura 7.10 Gráficas de funciones de densidad ji cuadrada

NOTACIÓN

Sea $\chi^2_{\alpha,\nu}$, llamado **valor crítico ji cuadrada,** el número sobre el eje horizontal de modo que α del área bajo la curva ji cuadrada con ν grados de libertad quede a la derecha de $\chi^2_{\alpha,\nu}$.

La simetría de las distribuciones t hizo que fuera necesario tabular sólo valores críticos t de cola superior ($t_{\alpha,\nu}$ con valores pequeños de α). La distribución ji cuadrada no es simétrica, por lo que la tabla A.7 del apéndice contiene valores de $\chi^2_{\alpha,\nu}$ tanto para α cerca de 0 como cerca de 1, como se ilustra en la figura 7.11(b). Por ejemplo, $\chi^2_{.025,14} = 26.119$, y $\chi^2_{.95,20}$ (el 5º percentil) $= 10.851$.

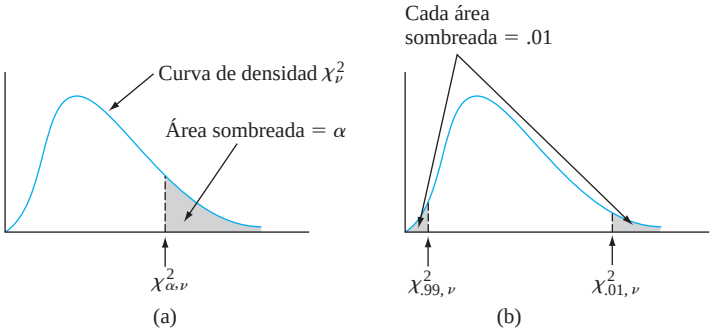


Figura 7.11 Notación $\chi^2_{\alpha,\nu}$ ilustrada

La variable aleatoria $(n-1)S^2/\sigma^2$ satisface las dos propiedades en las cuales está basado el método general para obtener un intervalo de confianza: es una función del parámetro de interés σ^2, no obstante su distribución de probabilidad (ji cuadrada) no depende de este parámetro. El área bajo una curva ji cuadrada con ν grados de libertad a la derecha de $\chi^2_{\alpha/2,\nu}$ es $\alpha/2$, lo mismo que a la izquierda de $\chi^2_{1-\alpha/2,\nu}$. De este modo el área capturada entre estos dos valores críticos es $1-\alpha$. Como una consecuencia de esto y del teorema que se acaba de formular,

$$P\left(\chi^2_{1-\alpha/2,n-1} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha/2,n-1}\right) = 1-\alpha \qquad (7.17)$$

Las desigualdades en (7.17) equivalen a

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}$$

Sustituyendo el valor calculado s^2 en los límites se obtiene un intervalo de confianza para σ^2, y sacando las raíces cuadradas se obtiene un intervalo para σ.

Un **intervalo de confianza de 100(1 − α)% para la varianza σ^2 de una población normal** tiene un límite inferior

$$(n-1)s^2/\chi^2_{\alpha/2,n-1}$$

y límite superior

$$(n-1)s^2/\chi^2_{1-\alpha/2,n-1}$$

Un **intervalo de confianza para** σ tiene límites superior e inferior que son las raíces cuadradas de los límites correspondientes en el intervalo para σ^2. Un límite de confianza superior o inferior resulta de la sustitución de $\alpha/2$ con α en el límite correspondiente del intervalo de confianza.

Ejemplo 7.15 Los datos adjuntos sobre voltaje de ruptura de circuitos eléctricamente sobrecargados se tomaron de un diagrama de probabilidad normal que apareció en el artículo "Damage of Flexible Printed Wiring Boards Associated with Lightning-Induced Voltage Surges", (*IEEE Transactions on Components, Hybrids, and Manuf. Tech.*, 1985: 214-220). La derechura del diagrama apoyó de manera firme la suposición de que el voltaje de ruptura está aproximadamente distribuido en forma normal.

1470	1510	1690	1740	1900	2000	2030	2100	2190
2200	2290	2380	2390	2480	2500	2580	2700	

Sea σ^2 la varianza de la distribución del voltaje de ruptura. El valor calculado de la varianza muestral es $s^2 = 137{,}324.3$, la estimación puntual de σ^2. Con grados de libertad $= n - 1 = 16$, un intervalo de confianza de 95% requiere $\chi^2_{.975,16} = 6.908$ y $\chi^2_{.025,16} = 28.845$. El intervalo es

$$\left(\frac{16(137{,}324.3)}{28.845}, \frac{16(137{,}324.3)}{6.908} \right) = (76{,}172.3, 318{,}064.4)$$

Sacando la raíz cuadrada de cada punto extremo se obtiene (276.0, 564.0) como el intervalo de confianza de 95% para σ. Estos intervalos son bastante anchos, lo que refleja la variabilidad sustancial del voltaje de ruptura en combinación con un tamaño de muestra pequeño. ■

Los intervalos de confianza para σ^2 y σ cuando la distribución de la población no es normal pueden ser difíciles de obtener. En esos casos, consulte a un estadístico conocedor.

EJERCICIOS Sección 7.4 (42–46)

42. Determine los valores de las siguientes cantidades:
- **a.** $\chi^2_{.1,15}$
- **b.** $\chi^2_{.1,25}$
- **c.** $\chi^2_{.01,25}$
- **d.** $\chi^2_{.005,25}$
- **e.** $\chi^2_{.99,25}$
- **f.** $\chi^2_{.995,25}$

43. Determine lo siguiente:
- **a.** El 95º percentil de la distribución ji cuadrada con $\nu = 10$
- **b.** El 5º percentil de la distribución ji cuadrada con $\nu = 10$.
- **c.** $P(10.98 \leq \chi^2 \leq 36.78)$, donde χ^2 es una variable aleatoria ji cuadrada con $\nu = 22$
- **d.** $P(\chi^2 < 14.611$ o $\chi^2 > 37.652)$, donde χ^2 es una variable aleatoria ji cuadrada con $\nu = 25$

44. Se determinó la cantidad de expansión lateral (mils) para una muestra de $n = 9$ soldaduras de arco de gas metálico de energía pulsante utilizadas en tanques de almacenamiento de buques LNG. La desviación estándar muestral resultante fue $s = 2.81$ mils. Suponiendo normalidad, obtenga un intervalo de confianza de 95% para σ^2 y para σ.

45. Se hicieron las siguientes observaciones de tenacidad a la fractura de una placa base de acero maraging con 18% de níquel ["Fracture Testing of Weldments", *ASTM Special Publ. No.* 381, 1965: 328-356 (en kg/pulg^2 $\sqrt{\text{in.}}$, dadas en orden creciente)]:

69.5	71.9	72.6	73.1	73.3	73.5	75.5	75.7
75.8	76.1	76.2	76.2	77.0	77.9	78.1	79.6
79.7	79.9	80.1	82.2	83.7	93.7		

Calcule un intervalo de confianza de 99% para la desviación estándar de la distribución de la tenacidad a la fractura. ¿Es válido este intervalo cualquiera que sea la naturaleza de la distribución? Explique.

46. El artículo "Concrete Pressure on Formwork" (*Mag. of Concrete Res.*, 2009: 407−417) dio las siguientes observaciones sobre la presión máxima del concreto (kN/m^2):

33.2	41.8	37.3	40.2	36.7	39.1	36.2	41.8
36.0	35.2	36.7	38.9	35.8	35.2	40.1	

- **a.** ¿Es factible que esta muestra fuera seleccionada de una distribución de población normal?
- **b.** Calcule un límite de confianza superior con nivel de confianza de 95% para la desviación estándar de la población de presión máxima.

EJERCICIOS SUPLEMENTARIOS (47–62)

47. El ejemplo 1.11 introdujo las observaciones adjuntas sobre fuerza de adhesión.

11.5	12.1	9.9	9.3	7.8	6.2	6.6	7.0
13.4	17.1	9.3	5.6	5.7	5.4	5.2	5.1
4.9	10.7	15.2	8.5	4.2	4.0	3.9	3.8
3.6	3.4	20.6	25.5	13.8	12.6	13.1	8.9
8.2	10.7	14.2	7.6	5.2	5.5	5.1	5.0
5.2	4.8	4.1	3.8	3.7	3.6	3.6	3.6

a. Calcule la fuerza de adhesión promedio verdadera de una manera que dé información sobre precisión y confiabilidad. [*Sugerencia*: $\Sigma x_i = 387.8$ y $\Sigma x_i^2 = 4247.08$.]

b. Calcule un intervalo de confianza de 95% para la proporción de todas las adhesiones cuyos valores de fuerza excederían de 10.

48. Un triatlón incluye natación, ciclismo y carrera a pie y es uno de los eventos deportivos amateurs más extenuantes. El artículo "Cardiovascular and Thermal Response of Triathlon Performance" (*Medicine and Science in Sports and Exercise*, 1988: 385–389) reporta sobre un estudio de investigación de nueve triatletas varones. Se registró el ritmo cardiaco máximo (pulsaciones/min) durante la actuación de cada uno de los tres eventos. Para natación, la media y la desviación estándar muestrales fueron 188.0 y 7.2, respectivamente. Suponiendo que la distribución de ritmo cardiaco es (de manera aproximada) normal, construya un intervalo de confianza de 98% para el ritmo cardiaco medio verdadero de triatletas mientras nadan.

49. Para cada uno de los 18 núcleos de depósitos de carbonato humedecidos con aceite, la cantidad de saturación de gas residual después de la inyección de un solvente se midió en la corriente de agua de salida. Las observaciones, en porcentaje de volumen de poros, fueron

23.5	31.5	34.0	46.7	45.6	32.5
41.4	37.2	42.5	46.9	51.5	36.4
44.5	35.7	33.5	39.3	22.0	51.2

(Véase "Relative Permeability Studies of Gas-Water Flow Following Solvent Injection in Carbonate Rocks", *Soc. of Petroleum Engineers J.*, 1976: 23–30.)

a. Construya una gráfica de caja de estos datos y comente sobre cualquier característica interesante.

b. ¿Es factible que la muestra fuera seleccionada de una distribución de población normal?

c. Calcule un intervalo de confianza de 98% para la cantidad promedio verdadera de saturación de gas residual.

50. Un artículo publicado en un periódico reporta que se utilizó una muestra de tamaño 5 como base para calcular un intervalo de confianza de 95% para la frecuencia natural (Hz) promedio verdadera de vigas deslaminadas de cierto tipo. El intervalo resultante fue (229.764, 233.504). Usted decide que un nivel de confianza de 99% es más apropiado que el de 95% utilizado. ¿Cuáles son los límites del intervalo de 99%? [*Sugerencia*: use el centro del intervalo y su ancho para determinar $\bar{x}$ y s.]

51. Una encuesta de 2253 adultos estadounidenses llevada a cabo por el Pew Research Center's Internet & American Life Project en el mes de abril del año 2009 reveló que 1262 de los encuestados había utilizado en algún momento medios inalámbricos para el acceso en línea.

a. Calcule e interprete un intervalo de confianza del 95% para la proporción de todos los adultos estadounidenses que en el momento de la encuesta habían usado medios inalámbricos para el acceso en línea.

b. ¿Qué tamaño de la muestra es necesario si la anchura deseada del intervalo de confianza del 95% debe ser como máximo .04, independientemente de los resultados de la muestra?

c. ¿El límite superior del intervalo en el inciso (a) especifica una fiabilidad del 95% en el límite superior para la proporción calculada? Explique.

52. La alta concentración del elemento tóxico arsénico es demasiado común en el agua subterránea. El artículo "Evaluation of Treatment Systems for the Removal of Arsenic from Groundwater" (*Practice Periodical of Hazardous, Toxic, and Radioactive Waste Mgmt.*, 2005: 152–157) reportó que para una muestra de $n = 5$ especímenes de agua seleccionada para tratamiento por coagulación, la concentración de arsénico media muestral fue de 24.3 μg/L y la desviación estándar muestral fue de 4.1. Los autores del artículo citado utilizaron métodos basados en t para analizar sus datos, así que venturosamente tuvieron razón al creer que la distribución de la concentración de arsénico era normal.

a. Calcule e interprete un intervalo de confianza de 95% para la concentración de arsénico promedio verdadera en todos los especímenes de agua.

b. Calcule un límite de confianza superior de 90% para la desviación estándar de la distribución de la concentración de arsénico.

c. Pronostique la concentración de arsénico de un solo espécimen de agua de modo que dé información sobre precisión y confiabilidad.

53. La infestación con pulgones de árboles frutales puede ser controlada rociando un pesticida o mediante el tratamiento con mariquitas. En un área particular, se seleccionan cuatro diferentes huertas de árboles frutales para experimentación. Las primeras tres arboledas se rocían con los pesticidas 1, 2 y 3, respectivamente, y la cuarta se trata con mariquitas con los siguientes resultados de cosecha:

Tratamiento	$n_i =$ Número de árboles	$\bar{x}_i$ (Bushels/árbol)	s_i
1	100	10.5	1.5
2	90	10.0	1.3
3	100	10.1	1.8
4	120	10.7	1.6

Sea μ_i = la cosecha promedio verdadera (bushels/árbol) después de recibir el i-ésimo tratamiento. Entonces

$$\theta = \frac{1}{3}(\mu_1 + \mu_2 + \mu_3) - \mu_4$$

mide la diferencia de las cosechas promedio verdaderas entre el tratamiento con pesticidas y el tratamiento con mariquitas. Cuando n_1, n_2, n_3 y n_4 son grandes, el estimador $\hat{\theta}$ obtenido al reemplazar cada μ_i con $\overline{X}_i$ es aproximadamente normal. Use esto para deducir un intervalo de confianza muestral grande de $100(1 - \alpha)\%$ para θ y calcule el intervalo de 95% para los datos dados.

54. Es importante que las máscaras utilizadas por bomberos sean capaces de soportar altas temperaturas porque los bomberos comúnmente trabajan en temperaturas de 200–500°F. En una prueba de un tipo de máscara, a 11 de 55 máscaras se les desprendió la mica a 250°. Construya un intervalo de confianza de 90% para la proporción verdadera de máscaras de este tipo cuya mica se desprendería a 250°.

55. Un fabricante de libros de texto universitarios está interesado en investigar la resistencia de las encuadernaciones producidas por una máquina de encuadernar particular. La resistencia puede ser medida registrando la fuerza requerida para arrancar las páginas de la encuadernación. Si esta fuerza se mide en libras, ¿cuántos libros deberán ser probados para calcular la fuerza promedio requerida para romper la encuadernación dentro de .1 lb con 95% de confianza? Suponga que se sabe que σ es de .8.

56. Es bien sabido que la exposición a la fibra de asbesto es un riesgo para la salud. El artículo "The Acute Effects of Chrysotile Asbestos Exposure on Lung Function" (*Environ. Research*, 1978: 360–372) reporta resultados sobre un estudio basado en una muestra de trabajadores de la construcción que habían estado expuestos a asbesto durante un periodo prolongado. Entre los datos dados en el artículo se encontraron los siguientes valores (ordenados) de elasticidad pulmonar ($cm^3/cm\ H_2O$) por cada uno de los 16 sujetos 8 meses después del periodo de exposición (la elasticidad pulmonar mide la elasticidad de los pulmones o cuán efectivamente son capaces de inhalar y exhalar):

167.9	180.8	184.8	189.8	194.8	200.2
201.9	206.9	207.2	208.4	226.3	227.7
228.5	232.4	239.8	258.6		

a. ¿Es factible que la distribución de la población sea normal?
b. Calcule un intervalo de confianza de 95% para la elasticidad pulmonar promedio verdadera después de la exposición.
c. Calcule un intervalo que, con un nivel de confianza de 95%, incluya por lo menos 95% de los valores de elasticidad pulmonar en la distribución de la población.

57. En el ejemplo 6.8, se introdujo el concepto de experimento censurado en el cual n componentes se prueban y el experimento termina en cuanto r de los componentes fallan. Suponga que las vidas útiles de los componentes son independientes, cada uno con distribución exponencial y parámetro λ. Sea Y_1 el tiempo en el cual ocurre la primera falla, Y_2 el tiempo en el cual ocurre la segunda falla, y así sucesivamente, de modo que $T_r = Y_1 + \cdots + Y_r + (n - r)Y_r$, es la vida útil total acumulada. En ese caso se puede demostrar que $2\lambda T_r$ tiene una distribución ji cuadrada con $2r$ grados de libertad. Use esto para desarrollar una fórmula para un intervalo de confianza de $100(1 - \alpha)\%$ para una vida útil promedio verdadera $1/\lambda$. Calcule un intervalo de confianza de 95% con los datos del ejemplo 6.8.

58. Sean $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución de probabilidad continua con mediana $\tilde{\mu}$ (de modo que $P(X_i \le \tilde{\mu}) = P(X_i \ge \tilde{\mu}) = .5$).

a. Demuestre que

$$P(\text{mín}\,(X_i) < \tilde{\mu} < \text{máx}\,(X_i)) = 1 - \left(\frac{1}{2}\right)^{n-1}$$

de modo que $(\text{mín}(x_i), \text{máx}(x_i))$ es un intervalo de confianza de $100(1 - \alpha)\%$ para $\tilde{\mu}$ con $\alpha = \left(\frac{1}{2}\right)^{n-1}$. [*Sugerencia*: el complemento del evento $\{\text{mín}\,(X_i) < \tilde{\mu} < \text{máx}\,(X_i)\}$ es $\{\text{máx}\,(X_i) \le \tilde{\mu}\} \cup \{\text{mín}\,(X_i) \ge \tilde{\mu}\}$. Pero $\text{máx}\,(X_i) \le \tilde{\mu}$ si y sólo si $X_i \le \tilde{\mu}$ con todas las i.]

b. Para cada uno de seis infantes normales varones, se determinó la cantidad de alanina aminoácida (mg/100 ml) mientras los infantes llevaban un dieta libre de isoleucina y se obtuvieron los siguientes resultados:

2.84 3.54 2.80 1.44 2.94 2.70

Calcule un intervalo de confianza de 97% para la cantidad mediana verdadera de alanina para infantes que llevaban esa dieta ("The Essential Amino Acid Requirements of Infants", *Amer. J. of Nutrition*, 1964: 322–330).

c. Sea $x_{(2)}$ la segunda más pequeña de las x_i y $x_{(n-1)}$ la segunda más grande de las x_i. ¿Cuál es el coeficiente de confianza del intervalo $(x_{(2)}, x_{(n-1)})$ para $\tilde{\mu}$?

59. Sean $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución uniforme en el intervalo $[0, \theta]$, de modo que

$$f(x) = \begin{cases} \frac{1}{\theta} & 0 \le x \le \theta \\ 0 & \text{de lo contrario} \end{cases}$$

Entonces si $Y = \text{máx}(X_i)$, se puede demostrar que la variable aleatoria $U = Y/\theta$ tiene una función de densidad

$$f_U(u) = \begin{cases} nu^{n-1} & 0 \le u \le 1 \\ 0 & \text{de lo contrario} \end{cases}$$

a. Use $f_U(u)$ para verificar que

$$P\left((\alpha/2)^{1/n} < \frac{Y}{\theta} \le (1 - \alpha/2)^{1/n}\right) = 1 - \alpha$$

y use esto para obtener un intervalo de confianza de $100(1 - \alpha)\%$ para θ.

b. Verifique que $P(\alpha^{1/n} \le Y/\theta \le 1) = 1 - \alpha$ y obtenga un intervalo de confianza de $100(1 - \alpha)\%$ para θ basado en esta proposición de probabilidad.

c. ¿Cuál de los dos intervalos obtenidos previamente es más corto? Si mi tiempo de espera de un camión en la mañana está uniformemente distribuido y los tiempos de espera observados son $x_1 = 4.2$, $x_2 = 3.5$, $x_3 = 1.7$, $x_4 = 1.2$ y $x_5 = 2.4$, obtenga un intervalo de confianza de 95% para θ utilizando el más corto de los dos intervalos.

60. Sea $0 \le \gamma \le \alpha$. Entonces un intervalo de confianza de $100(1 - \alpha)\%$ para μ cuando n es grande es

$$\left(\bar{x} - z_{\gamma} \cdot \frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha-\gamma} \cdot \frac{s}{\sqrt{n}}\right)$$

La opción $\gamma = \alpha/2$ da el intervalo usual obtenido en la sección 7.2; si $\gamma \neq \alpha/2$, este intervalo no es simétrico con respecto a $\bar{x}$. El ancho de este intervalo es $w = s(z_{\gamma} + z_{\alpha-\gamma})/\sqrt{n}$. Demuestre que w se reduce al mínimo con la opción $\gamma = \alpha/2$, de modo que el intervalo simétrico sea el más corto. [*Sugerencias*: (a) por definición de z_{α}, $\Phi(z_{\alpha}) = 1 - \alpha$, de modo que $z_{\alpha} = \Phi^{-1}(1 - \alpha)$: (b) la relación entre la derivada de una función $y = f(x)$ y la función inversa $x = f^{-1}(y)$ es $(d/dy)\, f^{-1}(y) = 1/f'(x)$.]

61. Suponga que $x_1, x_2, \ldots, x_n$ son valores observados resultantes de una muestra aleatoria tomada de una distribución simétrica pero posiblemente de cola gruesa. Sean $\tilde{x}$ y f_s la mediana muestral y la dispersión de los cuartos, respectivamente. El capítulo 11 de *Understanding Robust and Exploratory Data Analysis* (véase la bibliografía del capítulo 6) sugiere el siguiente intervalo de confianza de 95% robusto para la media de la población (punto de simetría):

$$\tilde{x} \pm \left(\frac{\text{valor crítico } t \text{ conservador}}{1.075}\right) \cdot \frac{f_s}{\sqrt{n}}$$

El valor de la cantidad entre paréntesis es 2.10 con $n = 10$, 1.94 con $n = 20$ y 1.91 con $n = 30$. Calcule este intervalo de confianza con los datos del ejercicio 45 y compare con el intervalo de confianza t apropiado para una distribución de población normal.

62. **a.** Use los resultados del ejemplo 7.5 para obtener un límite de confianza inferior de 95% para el parámetro λ de una distribución exponencial y calcule el límite basado en los datos dados en el ejemplo.

b. Si la vida útil tiene una distribución exponencial, la probabilidad de que la vida útil exceda de t es $P(X > t) = e^{-\lambda t}$. Use el resultado del inciso (a) para obtener un límite de confianza inferior de 95% para la probabilidad de que el tiempo de ruptura exceda de 100 min.

Bibliografía

DeGroot, Morris y Mark Schervish, *Probability and Statistics* (3a. ed.), Addison-Wesley, Reading, MA, 2002. Una muy buena exposición de los principios generales de inferencia estadística.

Devore, Jay y Kenneth Berk, *Modern Mathematical Statistics with Applications*, Cengage, Belmont, CA, 2007. La exposición es un poco más completa y sofisticada que la del presente libro e incluye más material sobre bootstrapping.

Hahn, Gerald y William Meeker, *Statistical Intervals*, Wiley, Nueva York, 1991. Todo lo que alguna vez quiso saber sobre intervalos estadísticos (de confianza, predicción, tolerancia y otros).

8 Pruebas de hipótesis basadas en una sola muestra

INTRODUCCIÓN

Un parámetro puede ser estimado a partir de datos muestrales con un solo número (una estimación puntual) o un intervalo completo de valores factibles (un intervalo de confianza). Con frecuencia, sin embargo, el objetivo de una investigación no es estimar un parámetro sino decidir cuál de dos afirmaciones contradictorias sobre el parámetro es la correcta. Los métodos para lograr esto comprenden la parte de la inferencia estadística llamada *prueba de hipótesis*. En este capítulo primero se discuten algunos de los conceptos y terminología básicos en la prueba de hipótesis y luego se desarrollan procedimientos para la toma de decisiones para los problemas de realización de pruebas con base en una muestra tomada de una sola población más frecuentemente encontrados.

8.1 Hipótesis y procedimientos de prueba

Una **hipótesis estadística** o simplemente *hipótesis* es una afirmación o aseveración sobre el valor de un solo parámetro (característica de una población o característica de una distribución de probabilidad), sobre los valores de varios parámetros o sobre la forma de una distribución de probabilidad completa. Un ejemplo de una hipótesis es la pretensión de que $\mu = .75$, donde μ es el diámetro interno promedio verdadero de un cierto tipo de tubo de PVC. Otro ejemplo es la proposición $p < .10$, donde p es la proporción de tarjetas de circuito defectuosas entre todas las tarjetas de circuito producidas por un fabricante. Si μ_1 y μ_2 denotan las resistencias a la ruptura promedio verdaderas de dos tipos diferentes de cuerdas, una hipótesis es la aseveración de que $\mu_1 - \mu_2 = 0$ y otra es que $\mu_1 - \mu_2 > 5$. No obstante, otro ejemplo de una hipótesis es la aseveración de que la distancia de detención en condiciones particulares tiene una distribución normal. Hipótesis de esta última clase se considerarán en el capítulo 14. En éste y en los siguientes capítulos, la atención se concentra en hipótesis en relación con parámetros.

En cualquier problema de prueba de hipótesis, existen dos hipótesis contradictorias en consideración. Una podría ser la pretensión de que $\mu = .75$ y la otra $\mu \neq .75$, o las dos proposiciones contradictorias podrían ser $p \geq .10$ y $p < .10$. El objetivo es decidir, con base en información muestral, cuál de las dos hipótesis es la correcta. Existe una analogía conocida de esto en un juicio criminal. Una pretensión es la aseveración de que el individuo acusado es inocente. En el sistema judicial estadounidense, ésta es la pretensión que inicialmente se cree que es cierta. Sólo de cara a una fuerte evidencia que diga lo contrario el jurado deberá rechazar esta pretensión a favor de la aseveración alternativa de que el acusado es culpable. En este sentido, la pretensión de inocencia es la hipótesis favorecida o protegida y la obligación de la comprobación recae en aquellos que creen en la pretensión alternativa.

Asimismo, al probar hipótesis estadísticas, el problema se formulará de modo que una de las pretensiones sea favorecida al inicio. Esta pretensión inicialmente favorecida no será rechazada a favor de la pretensión alternativa a menos que la evidencia muestral la contradiga y apoye con fuerza la aseveración alternativa.

DEFINICIÓN

La **hipótesis nula** denotada por H_0, es la pretensión que inicialmente se supone cierta (la pretensión de "creencia previa"). La **hipótesis alternativa** denotada por H_a, es la aseveración contradictoria de H_0.

La hipótesis nula será rechazada a favor de la hipótesis alternativa sólo si la evidencia muestral sugiere que H_0 es falsa. Si la muestra no contradice fuertemente a H_0, se continuará creyendo en la factibilidad de la hipótesis nula. Las dos posibles conclusiones derivadas de un análisis de prueba de hipótesis son entonces *rechazar* H_0 o *no rechazar* H_0.

Una **prueba de hipótesis** es un método de utilizar datos muestrales para decidir si la hipótesis nula debe ser rechazada. Por consiguiente se podría probar H_0; $\mu = .75$ contra la H_a alternativa: $\mu \neq .75$. Sólo si los datos muestrales sugieren fuertemente que μ es otra diferente de .75 deberá ser rechazada la hipótesis nula. Sin semejante evidencia, H_0 no deberá ser rechazada, puesto que sigue siendo bastante factible.

En ocasiones un investigador no desea aceptar una aseveración particular a menos y hasta que los datos apoyen fuertemente la aseveración. Como ejemplo, supóngase que una compañía está considerando aplicar un nuevo tipo de recubrimiento en los cojinetes que fabrica. Se sabe que la vida de desgaste promedio verdadera con el recubrimiento actual es de 1000 horas. Si μ denota la vida promedio verdadera del nuevo recubrimiento, la compañía no desea cambiar a menos que la evidencia sugiera fuertemente que μ excede de

1000. Una formulación apropiada del problema implicaría probar H_0: $\mu = 1000$ contra H_a: $\mu > 1000$. La conclusión de que se justifica un cambio está identificada con H_a y se requeriría evidencia conclusiva para justificar el rechazo de H_0 y cambiar al nuevo recubrimiento.

La investigación científica a menudo implica tratar de decidir si una teoría actual debe ser reemplazada por una explicación más factible y satisfactoria del fenómeno investigado. Un método conservador es identificar la teoría actual con H_0 y la explicación alternativa del investigador con H_a. El rechazo de la teoría actual ocurrirá entonces sólo cuando la evidencia sea mucho más compatible con la nueva teoría. En muchas situaciones, H_a se conoce como la "hipótesis del investigador", puesto que es la pretensión que al investigador en realidad le gustaría validar. La palabra *nula* significa "sin ningún valor, efecto o consecuencia", lo que sugiere que H_0 debería ser identificada con la hipótesis de ningún cambio (de la opinión actual), ninguna diferencia, ninguna mejora, y así sucesivamente. Supóngase, por ejemplo, que 10% de todas las tarjetas de circuito producidas por un fabricante durante un periodo reciente estaban defectuosas. Un ingeniero ha sugerido un cambio del proceso de producción en la creencia de que dará por resultado una proporción reducida de tarjetas defectuosas. Sea p la proporción verdadera de tarjetas defectuosas que resultan del proceso cambiado. Entonces la hipótesis de investigación en la cual recae la obligación de la comprobación, es la aseveración de que $p < .10$. Por consiguiente la hipótesis alternativa es H_a: $p < .10$.

En el tratamiento de la prueba de hipótesis, H_0 generalmente será formulada como pretensión de igualdad. Si θ denota el parámetro de interés, la hipótesis nula tendrá la forma H_0: $\theta = \theta_0$ donde θ_0 es un número específico llamado *valor nulo* del parámetro (valor pretendido para θ por la hipótesis nula). Como ejemplo, considérese la situación de la tarjeta de circuito que se acaba de discutir. La hipótesis alternativa sugerida fue H_a: $p < .10$, la pretensión de que la modificación del proceso reduce la proporción de tarjetas defectuosas. Una elección natural de H_0 en esta situación es la pretensión de que $p \geq .10$ de acuerdo con la cual el nuevo proceso no es mejor *o* peor que el actualmente utilizado. En su lugar se considerará H_0: $p = .10$ contra H_a: $p < .10$. El razonamiento para utilizar esta hipótesis nula simplificada es que cualquier procedimiento de decisión razonable para decidir entre H_0: $p = .10$ y H_a: $p < .10$ también será razonable para decidir entre la pretensión de que $p \geq .10$ y H_a. Se prefiere utilizar una H_0 simplificada porque tiene ciertos beneficios técnicos, los que en breve serán aparentes.

La alternativa a la hipótesis nula H_0: $\theta = \theta_0$ se verá como una de las siguientes tres aseveraciones:

1. H_a: $\theta > \theta_0$ (en cuyo caso la hipótesis nula implícita es $\theta \leq \theta_0$),

2. H_a: $\theta < \theta_0$ (en cuyo caso la hipótesis nula implícita es $\theta \geq \theta_0$), o

3. H_a: $\theta \neq \theta_0$

Por ejemplo, sea σ la desviación estándar de la distribución de diámetros internos (pulgadas) de cierto tipo de manguito de metal. Si se decidió utilizar el manguito a menos que la evidencia muestral demuestre conclusivamente que $\sigma > .001$, la hipótesis apropiada sería H_0: $\sigma = .001$ contra H_a: $\sigma > .001$. El número θ_0 que aparece tanto en H_0 como en H_a (separando la alternativa de la nula) se llama **valor nulo.**

Procedimientos de prueba

Un procedimiento de prueba es una regla, basada en datos muestrales, para decidir si se rechaza H_0. Una prueba de H_0: $p = .10$ contra H_a: $p < .10$ en el problema de tarjetas de circuito podría estar basada en examinar una muestra aleatoria de $n = 200$ tarjetas. Sea X el número de tarjetas defectuosas en la muestra, una variable aleatoria binomial; x representa el valor observado de X. Si H_0 es verdadera, $E(X) = np = 200(.10) = 20$, en tanto que se pueden esperar menos de 20 tarjetas defectuosas si H_a es verdadera. Un valor de x

un poco por debajo de 20 no contradice fuertemente a H_0, así que es razonable rechazar H_0 sólo si x es de manera sustancial menor que 20. Un procedimiento de prueba como ése es rechazar H_0 si $x \leq 15$ y no rechazarla de lo contrario. Este procedimiento consta de dos constituyentes: (1) un *estadístico de prueba* o función de los datos muestrales utilizados para tomar la decisión y (2) una *región de rechazo* compuesta de aquellos valores x con los cuales H_0 será rechazada a favor de H_a. De acuerdo con la regla que se acaba de sugerir, la región de rechazo se compone de $x = 0, 1, 2, \ldots,$ y 15. H_0 no será rechazada si $x = 16, 17, \ldots, 199$ o 200.

Un procedimiento de prueba se especifica como sigue:

1. Un **estadístico de prueba,** una función de los datos muestrales en los cuales ha de basarse la decisión (rechazar H_0 o no rechazar H_0)
2. Una **región de rechazo,** el conjunto de todos los valores estadísticos de prueba por los cuales H_0 será rechazada

La hipótesis nula será rechazada entonces si y sólo si el valor estadístico de prueba observado o calculado queda en la región de rechazo.

Como otro ejemplo, supóngase que una compañía tabacalera afirma que el contenido de nicotina promedio μ de los cigarrillos marca B es (cuando mucho) de 1.5 mg. Sería imprudente rechazar la afirmación del fabricante sin una fuerte evidencia contradictoria, así que una formulación apropiada del problema es probar H_0: $\mu = 1.5$ contra H_a: $\mu > 1.5$. Considérese una regla de decisión basada en analizar una muestra aleatoria de 32 cigarrillos. Sea $\overline{X}$ el contenido de nicotina promedio muestral. Si H_0 es verdadera $E(\overline{X}) = \mu = 1.5$ en tanto que si H_0 es falsa, se espera que $\overline{X}$ exceda de 1.5. Una fuerte evidencia contra H_0 es proporcionada por un valor de $\bar{x}$ que exceda considerablemente de 1.5. Por consiguiente, se podría utilizar $\overline{X}$ como un estadístico de prueba junto con la región de rechazo $\bar{x} \geq 1.6$.

Tanto en el ejemplo de tarjetas de circuito como en el de contenido de nicotina, la selección del estadístico de prueba y la forma de la región de rechazo tienen sentido intuitivamente. Sin embargo, la selección del valor de corte utilizado para especificar la región de rechazo es un tanto arbitraria. En lugar de rechazar H_0: $p = .10$ a favor de H_a: $p < .10$ cuando $x \leq 15$, se podría utilizar la región de rechazo $x \leq 14$. En esta región, H_0 no sería rechazada si se observaran 15 tarjetas defectuosas, mientras que esta ocurrencia conduciría al rechazo de H_0 si se emplea la región inicialmente sugerida. Asimismo, se podría utilizar la región de rechazo $\bar{x} \geq 1.55$ en el problema de contenido de nicotina en lugar de la región $\bar{x} \geq 1.60$.

Errores en la prueba de hipótesis

La base para elegir una región de rechazo particular radica en la consideración de los errores que se podrían presentar al sacar una conclusión. Considérese la región de rechazo $x \leq 15$ en el problema de tarjetas de circuito. Aun cuando H_0: $p = .10$ sea verdadera, podría suceder que una muestra inusual dé por resultado $x = 13$, de modo que H_0 sea erróneamente rechazada. Por otra parte, aun cuando H_a: $p < .10$ sea verdadera, una muestra inusual podría dar $x = 20$, en cuyo caso H_0 no sería rechazada, de nueva cuenta una conclusión incorrecta. Por lo tanto, es posible que H_0 pueda ser rechazada cuando sea verdadera o que H_0 no pueda ser rechazada cuando sea falsa. Estos posibles errores no son consecuencias de una región de rechazo imprudentemente seleccionada. Cualquiera de estos dos errores podría presentarse cuando se emplea la región $x \leq 14$, o cuando se utiliza cualquier otra región sensible.

DEFINICIÓN

Un **error de tipo I** consiste en rechazar la hipótesis nula H_0 cuando es verdadera.
Un **error de tipo II** implica no rechazar H_0 cuando H_0 es falsa.

En el problema de contenido de nicotina, un error de tipo I consiste en rechazar la afirmación del fabricante de que $\mu = 1.5$ cuando en realidad es cierta. Si se emplea la región de rechazo $\bar{x} \geq 1.6$, podría suceder que $\bar{x} = 1.63$ aun cuando $\mu = 1.5$, con el resultado de un error de tipo I. Alternativamente, puede ser que H_0 sea falsa y no obstante sea observada $\bar{x} = 1.52$, lo que conduciría a que H_0 no sea rechazada (un error de tipo II).

En el mejor de todos los mundos posibles, se podrían desarrollar procedimientos de prueba en los cuales ningún tipo de error es posible. Sin embargo, este ideal puede ser alcanzado sólo si la decisión se basa en el examen de toda la población. La dificultad con la utilización de un procedimiento basado en datos muestrales es que debido a la variabilidad del muestreo, el resultado podría ser una muestra no representativa, es decir, un valor de $\bar{X}$ que está lejos de μ o un valor de $\hat{p}$ que difiere considerablemente de p.

En lugar de demandar procedimientos sin errores, habrá que buscar procedimientos con los cuales sea improbable que ocurra cualquier tipo de error. Es decir, un buen procedimiento es uno con el cual la probabilidad de cometer cualquier tipo de error es pequeña. La selección de un valor de corte en una región de rechazo particular fija las probabilidades de errores de tipo I y tipo II. Estas probabilidades de error son tradicionalmente denotadas por α y β, de manera respectiva. Como H_0 especifica un valor único del parámetro, existe un solo valor de α. Sin embargo, existe un valor diferente de β para cada valor del parámetro compatible con H_a.

Ejemplo 8.1 Se sabe que cierto tipo de automóvil no sufre daños visibles el 25% del tiempo en pruebas de choque a 10 mph. Se ha propuesto un diseño de parachoques modificado en un esfuerzo por incrementar este porcentaje. Sea p la proporción de todos los choques a 10 mph con este nuevo parachoques en los que no se producen daños visibles. Las hipótesis a ser tratadas son H_0: $p = .25$ (ninguna mejora) contra H_a: $p > .25$. La prueba se basará en un experimento que implica $n = 20$ choques independientes con prototipos del nuevo diseño. Intuitivamente, H_0 deberá ser rechazada si un número sustancial de los choques no muestra daños. Considérese el siguiente procedimiento de prueba:

Estadístico de prueba: X = número de choques sin daños visibles
Región de rechazo: $R_8 = \{8, 9, 10, \ldots, 19, 20\}$; es decir, rechazar H_0 si $x \geq 8$, donde x es el valor observado del estadístico de prueba.

Esta región de rechazo se llama *de cola superior* porque se compone de sólo grandes valores del estadístico de prueba.

Cuando H_0 es verdadera, la distribución de probabilidad de X es binomial con $n = 20$ y $p = .25$. Entonces

$$\begin{aligned}\alpha &= P(\text{error de tipo I}) = P(H_0 \text{ es rechazada cuando es verdadera}) \\ &= P(X \geq 8 \text{ cuando } X \sim \text{Bin}(20, .25)) = 1 - B(7; 20, .25) \\ &= 1 - .898 = .102\end{aligned}$$

Es decir, cuando H_0 en realidad es verdadera, aproximadamente 10% de todos los experimentos compuestos de 20 choques darían por resultado que H_0 fuera rechazada incorrectamente (un error de tipo I).

En contraste con α, no hay una sola β. En su lugar, hay una β diferente por cada p distinta que exceda de .25. Por consiguiente, hay un valor de β con $p = .3$ (en cuyo caso $X \sim \text{Bin}(20, .3)$), otro valor de β con $p = .5$ y así sucesivamente. Por ejemplo,

$$\begin{aligned}\beta(.3) &= P(\text{error de tipo II cuando } p = .3) \\ &= P(H_0 \text{ no es rechazada cuando es falsa porque } p = .3) \\ &= P(X \leq 7 \text{ cuando } X \sim \text{Bin}(20, .3)) = B(7; 20, .3) = .772\end{aligned}$$

Cuando p es en realidad .3 y no .25 (un "pequeño" alejamiento de H_0), ¡aproximadamente 77% de todos los experimentos de este tipo darían por resultado que H_0 no sea rechazada de manera incorrecta!

La tabla adjunta muestra β para valores seleccionados de p (cada uno calculado para la región de rechazo R_8). Claramente, β disminuye conforme el valor de p se aleja hacia la derecha del valor nulo .25. De manera intuitiva, mientras más grande es el alejamiento de H_0, menos probable es que dicho alejamiento no sea detectado.

p	.3	.4	.5	.6	.7	.8
$\beta(p)$	.772	.416	.132	.021	.001	.000

El procedimiento de prueba propuesto sigue siendo razonable para poner a prueba la hipótesis nula más realista de que $p \leq .25$. En este caso, ya no existe una sola α, sino que hay una α por cada p que sea cuando mucho de .25: $\alpha(.25)$, $\alpha(.23)$, $\alpha(.20)$, $\alpha(.15)$ y así sucesivamente. Es fácil verificar, no obstante, que $\alpha(p) < \alpha(.25) = .102$ si $p < .25$. Es decir, el valor más grande de α ocurre con el valor límite .25 entre H_0 y H_a. Por consiguiente si α es pequeña para la hipótesis nula simplificada, también será igual o más pequeña para la H_0 más realista. ■

Ejemplo 8.2 Se sabe que el tiempo de secado de un tipo de pintura en condiciones de prueba especificadas está normalmente distribuido con valor medio de 75 min y desviación estándar de 9 min. Algunos químicos propusieron un nuevo aditivo para reducir el promedio de tiempo de secado. Se cree que los tiempos de secado con este aditivo permanecerán distribuidos en forma normal con $\sigma = 9$. Debido al gasto asociado con el aditivo, la evidencia deberá sugerir fuertemente una mejora en el tiempo de secado promedio antes de que se adopte semejante conclusión. Sea μ el tiempo de secado promedio verdadero cuando se utiliza el aditivo. Las hipótesis apropiadas son H_0: $\mu = 75$ contra H_a: $\mu < 75$. Sólo si H_0 puede ser rechazada el aditivo será declarado exitoso y utilizado.

Los datos experimentales tienen que estar compuestos de tiempos de secado de $n = 25$ especímenes de prueba. Sean $X_1, \ldots, X_{25}$ los 25 tiempos de secado, una muestra aleatoria de tamaño 25 de una distribución normal con valor medio μ y desviación estándar $\sigma = 9$. El tiempo de secado medio muestral $\overline{X}$ tiene entonces una distribución normal con valor esperado $\mu_{\overline{X}} = \mu$ y desviación estándar $\sigma_{\overline{X}} = \sigma/\sqrt{n} = 9/\sqrt{25} = 1.80$. Cuando H_0 es verdadera, $\mu_{\overline{X}} = 75$, así que un valor $\overline{x}$ sustancialmente menor que 75 contradiría fuertemente a H_0. Una región razonable de rechazo tiene la forma $\overline{x} \leq c$, donde el valor de corte c es adecuadamente seleccionado. Considere la opción $c = 70.8$, de modo que el procedimiento de prueba se componga del estadístico de prueba $\overline{X}$ y una región de rechazo $\overline{x} \leq 70.8$. Debido a que la región de rechazo se compone de sólo valores pequeños del estadístico de prueba, se dice que ésta es *de cola pequeña*. El cálculo de α y β ahora implica una estandarización de rutina de $\overline{X}$ seguida por una referencia a las probabilidades normales estándar de la tabla A.3 del apéndice:

$$\begin{aligned}\alpha &= P(\text{error de tipo I}) = P(H_0 \text{ es rechazada cuando es verdadera}) \\ &= P(\overline{X} \leq 70.8 \text{ cuando } \overline{X} \sim \text{normal con } \mu_{\overline{X}} = 75, \sigma_{\overline{X}} = 1.8) \\ &= \Phi\left(\frac{70.8 - 75}{1.8}\right) = \Phi(-2.33) = .01\end{aligned}$$

$$\begin{aligned}\beta(72) &= P(\text{error de tipo II cuando } \mu = 72) \\ &= P(H_0 \text{ no es rechazada cuando es falsa porque } \mu = 72) \\ &= P(\overline{X} > 70.8 \text{ cuando } \overline{X} \sim \text{normal con } \mu_{\overline{X}} = 72 \text{ y } \sigma_{\overline{X}} = 1.8) \\ &= 1 - \Phi\left(\frac{70.8 - 72}{1.8}\right) = 1 - \Phi(-.67) = 1 - .2514 = .7486\end{aligned}$$

$$\beta(70) = 1 - \Phi\left(\frac{70.8 - 70}{1.8}\right) = .3300 \qquad \beta(67) = .0174$$

Para el procedimiento de prueba especificado, sólo 1% de todos los experimentos realizados como se describió darán por resultado que H_0 sea rechazada cuando en realidad es verdadera. No obstante, la probabilidad de un error de tipo II es muy grande cuando $\mu = 72$ (sólo un pequeño alejamiento de H_0), un poco menor cuando $\mu = 70$ y bastante pequeño cuando $\mu = 67$ (un alejamiento muy sustancial de H_0). Estas probabilidades de error se ilustran en la figura 8.1. Obsérvese que α se calcula con la distribución de probabilidad del estadístico de prueba cuando H_0 es verdadera, en tanto que la determinación de β requiere conocer la distribución del estadístico de prueba cuando H_0 es falsa.

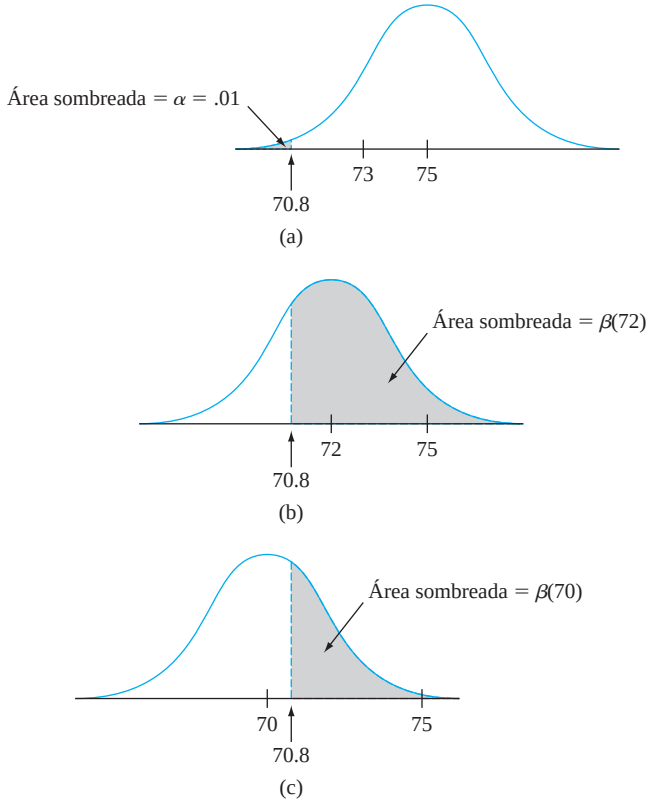


Figura 8.1 α y β ilustradas para el ejemplo 8.2: (a) la distribución de $\bar{X}$ cuando $\mu = 75$ (H_0 verdadera); (b) la distribución de $\bar{X}$ cuando $\mu = 72$ (H_0 falsa); (c) la distribución de $\bar{X}$ cuando $\mu = 70$ (H_0 falsa)

Como en el ejemplo 8.1, si se considera la hipótesis nula más realista $\mu \geq 75$, existe una α para cada valor de parámetro con el cual H_0 es verdadera: $\alpha(75)$, $\alpha(75.8)$, $\alpha(76.5)$, y así de manera sucesiva. Es fácil verificar que $\alpha(75)$ es la más grande de todas estas probabilidades de error de tipo I. Enfocar en el valor límite equivale a trabajar explícitamente con el "peor caso". ■

La especificación de un valor de corte para la región de rechazo en los ejemplos que se acaban de considerar fue algo arbitraria. El uso de $R_8\{8, 9, \ldots, 20\}$ en el ejemplo 8.1 dio por resultado $\alpha = .102$, $\beta(.3) = .772$ y $\beta(.5) = .132$. Muchos pensarán que estas probabilidades de error son intolerablemente grandes. Quizás puedan reducirse si se cambia el valor de corte.

Ejemplo 8.3 (Continuación del ejemplo 8.1)

Utilice el mismo experimento y el estadístico de prueba X como previamente se describió en el problema de la defensa de automóvil pero ahora considere la región de rechazo $R_9 = \{9, 10, \ldots, 20\}$. Como X sigue teniendo una distribución binomial con parámetros $n = 20$ y p,

$$\begin{aligned}\alpha &= P(H_0 \text{ es rechazada cuando } p = .25)\\ &= P(X \geq 9 \text{ cuando } X \sim \text{Bin}(20, .25)) = 1 - B(8; 20, .25) = .041\end{aligned}$$

La probabilidad de error de tipo I se redujo con el uso de la nueva región de rechazo. Sin embargo, se pagó un precio por esta reducción:

$$\begin{aligned}\beta(.3) &= P(H_0 \text{ no es rechazada cuando } p = .3)\\ &= P(X \le 8 \text{ cuando } X \sim \text{Bin}(20, .3)) = B(8; 20, .3) = .887\\ \beta(.5) &= B(8; 20, .5) = .252\end{aligned}$$

Ambas β son más grandes que las probabilidades de error correspondientes .772 y .132 para la región R_8. En retrospectiva, esto no es sorprendente; α se calcula sumando las probabilidades de los valores estadísticos de prueba *en la región de rechazo*, en tanto que β es la probabilidad de que X quede *en el complemento* de la región de rechazo. Al hacerse más pequeña la región de rechazo debe reducirse α al mismo tiempo que se incrementa β con cualquier $p > .25$. ■

Ejemplo 8.4 (Continuación del ejemplo 8.2)

El uso del valor de corte $c = 70.8$ en el ejemplo de secado de la pintura dio por resultado un valor de α muy pequeño (.01) pero β grandes. Considere el mismo experimento y el estadístico de prueba $\bar{X}$ con la nueva región de rechazo $\bar{x} \le 72$. Como $\bar{X}$ sigue siendo normalmente distribuida con valor medio $\mu_{\bar{X}} = \mu$ y $\sigma_{\bar{X}} = 1.8$,

$$\begin{aligned}\alpha &= P(H_0 \text{ es rechazada cuando es verdadera})\\ &= P(\bar{X} \le 72 \text{ cuando } \bar{X} \sim N(75, 1.8^2))\\ &= \Phi\left(\frac{72 - 75}{1.8}\right) = \Phi(-1.67) = .0475 \approx .05\end{aligned}$$

$$\begin{aligned}\beta(72) &= P(H_0 \text{ no es rechazada cuando } \mu = 72)\\ &= P(\bar{X} > 72 \text{ cuando } \bar{X} \text{ es una variable aleatoria normal con media de 72 y desviación estándar de 1.8})\\ &= 1 - \Phi\left(\frac{72 - 72}{1.8}\right) = 1 - \Phi(0) = .5\end{aligned}$$

$$\beta(70) = 1 - \Phi\left(\frac{72 - 70}{1.8}\right) = .1335 \qquad \beta(67) = .0027$$

El cambio del valor de corte agrandó la región de rechazo (incluye más valores de $\bar{x}$) y el resultado es una reducción de β por cada μ fija menor que 75. Sin embargo, α en esta nueva región se ha incrementado desde el valor previo .01 hasta aproximadamente .05. Si una probabilidad de error de tipo I así de grande puede ser tolerada, se prefiere la segunda región ($c = 72$) a la primera ($c = 70.8$) debido a las β más pequeñas. ■

Los resultados de estos ejemplos pueden ser generalizados de la siguiente manera.

PROPOSICIÓN

Supóngase que un experimento y un tamaño de muestra están fijos y que se selecciona un estadístico de prueba. Entonces si se reduce el tamaño de la región de rechazo para obtener un valor más pequeño de α se obtiene un valor más grande de β con cualquier valor de parámetro particular compatible con H_a.

Esta proposición expresa que una vez que el estadístico de prueba y n están fijos, no existe una región de rechazo que haga que al mismo tiempo α y β sean pequeños. Se debe seleccionar una región para establecer un compromiso entre α y β.

Debido a las indicaciones sugeridas para especificar H_0 y H_a, casi siempre un error de tipo I es más serio que uno de tipo II (esto en general se puede lograr mediante la selección apropiada de las hipótesis). El método seguido por la mayoría de los practicantes de la estadística es especificar el valor más grande de α que pueda ser tolerado y encontrar una región de rechazo que tenga ese valor de α en lugar de cualquier otro más pequeño. Esto hace a β tan pequeño como sea posible sujeto al límite en α. El valor resultante de α

a menudo se conoce como **nivel de significancia** de la prueba. Niveles tradicionales de significancia son .10, .05 y .01, aunque el nivel en cualquier problema particular dependerá de la seriedad de un error de tipo I; mientras más serio es este error, más pequeño deberá ser el nivel de significancia. El procedimiento de prueba correspondiente se llama **prueba de nivel α** (p. ej., prueba de nivel .05 o prueba de nivel .01). Una prueba con nivel de significancia α es una donde la probabilidad de error de tipo I se controla al nivel especificado.

Ejemplo 8.5 Nuevamente sea μ que denota el contenido de nicotina promedio verdadero de los cigarrillos marca B. El objetivo es probar H_0: $\mu = 1.5$ contra H_a: $\mu > 1.5$ con base en una muestra aleatoria $X_1, X_2, \ldots, X_{32}$ de contenido de nicotina. Suponga que se sabe que la distribución del contenido de nicotina es normal con $\sigma = .20$. Entonces $\overline{X}$ está normalmente distribuida con valor medio $\mu_{\overline{X}} = \mu$ y desviación estándar $\sigma_{\overline{X}} = .20/\sqrt{32} = .0354$.

En lugar de utilizar $\overline{X}$ como estadístico de prueba, se estandariza $\overline{X}$ suponiendo que H_0 es verdadera.

$$\text{Estadístico de prueba: } Z = \frac{\overline{X} - 1.5}{\sigma/\sqrt{n}} = \frac{\overline{X} - 1.5}{.0354}$$

Z expresa la distancia entre $\overline{X}$ y su valor esperado cuando H_0 es verdadera como algún número de desviaciones estándar. Por ejemplo, $z = 3$ resulta de una $\bar{x}$ que es 3 desviaciones estándar más grande de lo que se habría esperado si H_0 fuera verdadera.

Rechazar H_0 cuando $\bar{x}$ excede "considerablemente" de 1.5 equivale a rechazar H_0 cuando z excede "considerablemente" de cero. Es decir, la forma de la región de rechazo es $z \geq c$. Determínese ahora c de modo que $\alpha = .05$. Cuando H_0 es verdadera, Z tiene una distribución estándar normal. Por consiguiente

$$\begin{aligned}\alpha &= P(\text{error de tipo I}) = P(\text{rechazar } H_0 \text{ cuando } H_0 \text{ es verdadera})\\ &= P(Z \geq c \text{ cuando } Z \sim N(0, 1))\end{aligned}$$

El valor de c debe capturar el área de la cola superior .05 bajo la curva z. De la sección 4.3 o directamente de la tabla A.3 del apéndice, $c = z_{.05} = 1.645$.

Obsérvese que $z \geq 1.645$ equivale a $\bar{x} - 1.5 \geq (.0354)(1.645)$, es decir, $\bar{x} \geq 1.56$. Entonces β implica la probabilidad de que $\overline{X} < 1.56$ y puede ser calculada para cualquier μ mayor que 1.5. ■

EJERCICIOS Sección 8.1 (1–14)

1. Por cada una de las siguientes aseveraciones, exprese si es una hipótesis estadística legítima y por qué:
- **a.** H: $\sigma > 100$
- **b.** H: $\tilde{x} = 45$
- **c.** H: $s \leq .20$
- **d.** H: $\sigma_1/\sigma_2 < 1$
- **e.** H: $\overline{X} - \overline{Y} = 5$
- **f.** H: $\lambda \leq .01$ donde λ es el parámetro de una distribución exponencial utilizada para modelar la vida útil de un componente

2. Para los siguientes pares de aseveraciones, indique cuáles no satisfacen las reglas para establecer hipótesis y por qué (los subíndices 1 y 2 diferencian las cantidades para dos poblaciones o muestras diferentes).
- **a.** H_0: $\mu = 100$, H_a: $\mu > 100$
- **b.** H_0: $\sigma = 20$, H_a: $\sigma \leq 20$
- **c.** H_0: $p \neq .25$, H_a: $p = .25$
- **d.** H_0: $\mu_1 - \mu_2 = 25$, H_a: $\mu_1 - \mu_2 > 100$
- **e.** H_0: $S_1^2 = S_2^2$, H_a: $S_1^2 \neq S_2^2$
- **f.** H_0: $\mu = 120$, H_a: $\mu = 150$
- **g.** H_0: $\sigma_1/\sigma_2 = 1$, H_a: $\sigma_1/\sigma_2 \neq 1$
- **h.** H_0: $p_1 - p_2 = -.1$, H_a: $p_1 - p_2 < -.1$

3. Para determinar si las soldaduras de las tuberías en una planta de energía nuclear satisfacen las especificaciones, se selecciona una muestra aleatoria de soldaduras y se realizan pruebas en cada una de ellas. La resistencia de la soldadura se mide como la fuerza requerida para romperla. Suponga que las especificaciones indican que la resistencia media de las soldaduras deberá exceder de 100 lb/pulg^2; el equipo de inspección decide probar H_0: $\mu = 100$ contra H_a: $\mu > 100$. Explique por qué podría ser preferible utilizar esta H_a en lugar de $\mu < 100$.

4. Sea μ el nivel de radiactividad promedio verdadero (picocuries por litro). Se considera que el valor 5 pCi/L es la línea divisoria entre agua segura e insegura. ¿Recomendaría probar H_0: $\mu = 5$ contra H_a: $\mu > 5$ o H_0: $\mu = 5$ contra H_a: $\mu < 5$? Explique su razonamiento [*Sugerencia*: piense en las consecuencias de un error de tipo I o de un error de tipo II con cada posibilidad.]

5. Antes de aprobar la compra de un gran pedido de fundas de polietileno para un tipo particular de cable de energía submarino relleno de aceite a alta presión, una compañía desea contar con evidencia conclusiva de que la desviación estándar verdadera del espesor de la funda es de menos de .05 mm. ¿Qué hipótesis deberán ser probadas y por qué? En este contexto, ¿cuáles son los errores de tipos I y II?

6. Muchas casas viejas cuentan con sistemas eléctricos que utilizan fusibles en lugar de interruptores de circuito. Un fabricante de fusibles de 40 amp desea asegurarse de que el amperaje medio al cual se queman sus fusibles es en realidad de 40. Si el amperaje medio es menor que 40, los clientes se quejarán porque los fusibles tienen que ser reemplazados con demasiada frecuencia. Si el amperaje medio es de más de 40, el fabricante podría ser responsable de los daños que sufra un sistema eléctrico a causa del funcionamiento defectuoso de los fusibles. Para verificar el amperaje de los fusibles, se selecciona e inspecciona una muestra de fusibles. Si tuviera que realizarse un prueba de hipótesis con los datos resultantes, ¿qué hipótesis nula y alternativa serían de interés para el fabricante? Describa los errores de tipos I y II en el contexto de este problema.

7. Se toman muestras de agua utilizada para enfriamiento al momento de ser descargada por una planta de energía en un río. Se ha determinado que en tanto la temperatura media del agua descargada sea cuando mucho de 150°F, no habrá efectos negativos en el ecosistema del río. Para investigar si la planta cumple con los reglamentos que prohíben una temperatura media por encima de 150° del agua de descarga, se tomarán al azar 50 muestras de agua y se registrará la temperatura de cada una. Los datos resultantes se utilizarán para probar la hipótesis H_0: $\mu = 150°$ contra H_a: $\mu > 150°$. En el contexto de esta situación, describa los errores de tipo I y tipo II. ¿Qué tipo de error consideraría más serio? Explique.

8. Un tipo regular de laminado está siendo utilizado por un fabricante de tarjetas de circuito. Un laminado especial ha sido desarrollado para reducir la combadura. El laminado regular será utilizado en una muestra de especímenes y el laminado especial en otra muestra y se determinará entonces la cantidad de combadura en cada espécimen. El fabricante cambiará entonces al laminado especial sólo si puede demostrar que la cantidad de combadura promedio verdadera de dicho laminado es menor que la del laminado regular. Formule las hipótesis pertinentes y describa los errores de tipo I y de tipo II en el contexto de esta situación.

9. Dos compañías diferentes han solicitado proporcionar el servicio de televisión por cable en una región. Sea p la proporción de todos los suscriptores potenciales que favorecen a la primera compañía sobre la segunda. Considere probar H_0: $p = .5$ contra H_a: $p \neq .5$ basado en una muestra aleatoria de 25 individuos. Sea X el número en la muestra que favorecen a la primera compañía y x el valor observado de X.

a. ¿Cuál de las siguientes regiones de rechazo es más apropiada y por qué?

$$R_1 = \{x: x \leq 7 \text{ o } x \geq 18\}, R_2 = \{x: x \leq 8\},$$
$$R_3 = \{x: x \geq 17\}$$

b. En el contexto de este problema, describa cuáles son los errores de tipo I y de tipo II.

c. ¿Cuál es la distribución de probabilidad del estadístico de prueba X cuando H_0 es verdadera? Úsela para calcular la probabilidad de un error de tipo I.

d. Calcule la probabilidad de un error de tipo II en la región seleccionada cuando $p = .3$, otra vez cuando $p = .4$ y también con $p = .6$ y $p = .7$.

e. Utilizando la región seleccionada, ¿qué concluiría si 6 de los 25 individuos encuestados favorecen a la compañía 1?

10. Una mezcla de cenizas combustibles pulverizadas y cemento Portland utilizada para rellenar con lechada deberá tener una resistencia a la compresión de más de 1300 KN/m^2. La mezcla no será utilizada a menos que la evidencia experimental indique concluyentemente que la especificación de resistencia ha sido satisfecha. Suponga que la resistencia a la compresión de especímenes de esta muestra está normalmente distribuida con $\sigma = 60$. Sea μ la resistencia a la compresión promedio verdadera.

a. ¿Cuáles son las hipótesis nula y alternativa apropiadas?

b. Sea $\overline{X}$ la resistencia a la compresión promedio muestral de $n = 10$ especímenes seleccionados al azar. Considere el procedimiento de prueba con estadístico de prueba $\overline{X}$ y región de rechazo $\overline{x} \geq 1331.26$. ¿Cuál es la distribución de probabilidad del estadístico de prueba cuando H_0 es verdadera? ¿Cuál es la probabilidad de un error de tipo I para el procedimiento de prueba?

c. ¿Cuál es la distribución de probabilidad del estadístico de prueba cuando $\mu = 1350$? Utilizando el procedimiento de prueba del inciso (b), ¿cuál es la probabilidad de que la mezcla sea juzgada insatisfactoria cuando en realidad $\mu = 1350$ (un error de tipo II)?

d. ¿Cómo cambiaría el procedimiento de prueba del inciso (b) para obtener una prueba con nivel de significancia de .05? ¿Qué impacto tendría este cambio en la probabilidad de error del inciso (c)?

e. Considere el estadístico de prueba estandarizado $Z = (\overline{X} - 1300)/(\sigma/\sqrt{n}) = (\overline{X} - 1300)/13.42$. ¿Cuáles son los valores de Z correspondientes a la región de rechazo del inciso (b)?

11. La calibración de una báscula tiene que ser verificada pesando 25 veces un espécimen de prueba de 10 kg. Suponga que los resultados de diferentes pesadas son independientes entre sí y que el peso en cada ensayo está normalmente distribuido con $\sigma = .200$ kg. Sea μ la lectura de peso promedio verdadero en la báscula.

a. ¿Qué hipótesis deberá ponerse a prueba?

b. Suponga que la báscula tiene que ser recalibrada si $\overline{x} \geq 10.1032$ o $\overline{x} \leq 9.8968$. ¿Cuál es la probabilidad de que se realice la recalibración cuando en realidad no es necesaria?

c. ¿Cuál es la probabilidad de que la recalibración sea considerada innecesaria cuando en realidad $\mu = 10.1$? ¿Cuando $\mu = 9.8$?

d. Sea $z = (\overline{x} - 10)/(\sigma/\sqrt{n})$. ¿Con qué valor de c la región de rechazo del inciso (b) equivale a la región de "dos colas" $z \geq c$ o $z \leq -c$?

e. Si el tamaño de muestra fuera de sólo 10 y no de 25, ¿cómo modificaría el procedimiento del inciso (d) de modo que $\alpha = .05$?

f. Utilizando la prueba del inciso (e), ¿qué concluiría basado en los siguientes datos muestrales?

9.981	10.006	9.857	10.107	9.888
9.728	10.439	10.214	10.190	9.793

g. Exprese de nuevo el procedimiento de prueba del inciso (b) en función del estadístico de prueba estandarizado $Z = (\bar{X} - 10)/(\sigma/\sqrt{n})$.

12. Se ha propuesto un nuevo diseño del sistema de frenos de un tipo de carro. Para el sistema actual, se sabe que la distancia de frenado promedio verdadera a 40 mph en condiciones específicas es de 120 pies. Se propone que el nuevo diseño sea instalado sólo si los datos muestrales indican fuertemente una reducción de la distancia de frenado promedio verdadera del nuevo diseño.

a. Defina el parámetro de interés y formule las hipótesis pertinentes.

b. Suponga que la distancia de frenado del nuevo sistema está normalmente distribuida con $\sigma = 10$. Sea $\bar{X}$ la distancia de frenado promedio de una muestra aleatoria de 36 observaciones. ¿Cuál de las siguientes tres regiones de rechazo es apropiada: $R_1 = \{\bar{x}: \bar{x} \geq 124.80\}$, $R_2 = \{\bar{x}: \bar{x} \leq 115.20\}$, $R_3 = \{\bar{x}: \bar{x} \geq 125.13 \text{ o } \bar{x} \leq 114.87\}$?

c. ¿Cuál es el nivel de significancia de la región apropiada del inciso (b)? ¿Cómo cambiaría la región para obtener una prueba con $\alpha = .001$?

d. ¿Cuál es la probabilidad de que el nuevo diseño no sea instalado cuando la distancia de frenado promedio verdadera sea en realidad de 115 pies y la región apropiada del inciso (b) sea utilizada?

e. Sea $Z = (\bar{X} - 120)/(\sigma/\sqrt{n})$. ¿Cuál es el nivel de significancia de la región de rechazo $\{z: z \leq -2.33\}$? ¿Y para la región $\{z: z \leq -2.88\}$?

13. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución de población normal con un valor conocido de σ.

a. Para probar las hipótesis H_0: $\mu = \mu_0$, contra H_a: $\mu > \mu_0$ (donde μ_0 es un número fijo), demuestre que la prueba con el estadístico de prueba $\bar{X}$ y región de rechazo $\bar{x} \geq \mu_0 + 2.33\sigma/\sqrt{n}$ tiene un nivel de significancia de .01.

b. Suponga que se utiliza el procedimiento del inciso (a) para probar H_0: $\mu \leq \mu_0$ contra H_a: $\mu > \mu_0$. Si $\mu_0 = 100$, $n = 25$ y $\sigma = 5$, ¿cuál es la probabilidad de cometer un error de tipo I cuando $\mu = 99$? ¿Cuando $\mu = 98$? En general, ¿qué se puede decir sobre la probabilidad de un error de tipo I cuando el valor real de μ es menor que μ_0? Verifique su aseveración.

14. Reconsidere la situación del ejercicio 11 y suponga que la región de rechazo es $\{\bar{x}: \bar{x} \geq 10.1004 \text{ o } \bar{x} \leq 9.8940\} = \{z: z \geq 2.51 \text{ o } z \leq -2.65\}$.

a. ¿Cuál es α para este procedimiento?

b. ¿Cuál es β cuando $\mu = 10.1$? ¿Cuando $\mu = 9.9$? ¿Es ésta deseable?

8.2 Pruebas sobre una media de población

La discusión general en el capítulo 7 de intervalos de confianza para una media de población μ se enfocó en tres casos diferentes. A continuación se desarrollan procedimientos de prueba para estos casos.

Caso I: Una población normal con σ conocida

Aun cuando la suposición de que el valor de σ es conocido rara vez se cumple en la práctica, este caso proporciona un buen punto de partida debido a la facilidad con la que los procedimientos generales y sus propiedades pueden ser desarrollados. La hipótesis nula en los tres casos propondrá que μ tiene un valor numérico particular, el *valor nulo*, el cual será denotado por μ_0. Sea $X_1, \ldots, X_n$ una muestra aleatoria de tamaño n de la población normal. Entonces la media muestral $\bar{X}$ tiene una distribución normal con valor esperado $\mu_{\bar{X}} = \mu$ y desviación estándar $\sigma_{\bar{X}} = \sigma/\sqrt{n}$. Cuando H_0 es verdadera $\mu_{\bar{X}} = \mu_0$. Considérese ahora el estadístico Z obtenido estandarizando $\bar{X}$ dada la suposición de que H_0 es verdadera:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

Al sustituir la media muestral calculada $\bar{x}$ se obtiene z, la distancia entre $\bar{x}$ y μ_0 expresada en "unidades de desviación estándar". Por ejemplo, si la hipótesis nula es H_0: $\mu = 100$, $\sigma_{\bar{X}} = \sigma/\sqrt{n} = 10/\sqrt{25} = 2.0$ y $\bar{x} = 103$, entonces el valor estadístico de prueba es $z = (103 - 100)/2.0 = 1.5$. Es decir, el valor observado de $\bar{x}$ es 1.5 desviaciones estándar (de $\bar{X}$) más grande de lo que se espera que sea cuando H_0 es verdadera. El estadístico Z es una medida natural de la distancia entre $\bar{X}$, el estimador de μ, y su valor esperado cuando H_0 es verdadera. Si esta distancia es demasiado grande en una dirección consistente con H_a, la hipótesis nula deberá ser rechazada.

Supóngase primero que la hipótesis alternativa tiene la forma H_a: $\mu > \mu_0$. Entonces un valor de $\bar{x}$ menor que μ_0 indudablemente no apoya a H_a. Tal $\bar{x}$ corresponde a un valor negativo de z (puesto que $\bar{x} - \mu_0$ es negativo y el divisor de $\sigma/\sqrt{n}$ es positivo). Del mismo modo, un valor de $\bar{x}$ que exceda de μ_0 por sólo una pequeña cantidad (correspondiente a z la cual es positiva aunque pequeña) no sugiere que H_0 deberá ser rechazada a favor de H_a. El rechazo de H_0 es apropiado sólo cuando $\bar{x}$ excede considerablemente de μ_0; es decir, cuando el valor de z es positivo y grande. En suma, la región de rechazo apropiada basada en el estadístico de prueba Z en lugar de $\bar{X}$ tiene la forma $z \geq c$.

Como se discutió en la sección 8.1, el valor de corte c deberá ser elegido para controlar la probabilidad de un error de tipo I al nivel α deseado. Esto es fácil de lograr porque la distribución del estadístico de prueba Z cuando H_0 es verdadera es la distribución normal estándar (es por eso que μ_0 se restó al estandarizar). El valor c de corte requerido es el valor crítico z que captura el área de la cola superior α bajo la curva z. Como ejemplo, sea $c = 1.645$, el valor que captura el área de cola .05 ($z_{.05} = 1.645$). Entonces,

$$\begin{aligned}\alpha &= P(\text{error de tipo I}) = P(H_0 \text{ es rechazada cuando } H_0 \text{ es verdadera})\\ &= P(Z \geq 1.645 \text{ cuando } Z \sim N(0,1)) = 1 - \Phi(1.645) = .05\end{aligned}$$

Más generalmente, la región de rechazo $z \geq z_\alpha$ tiene una probabilidad α de error de tipo I. El procedimiento de prueba es de *cola superior* porque la región de rechazo se compone sólo de valores grandes del estadístico de prueba.

Un razonamiento análogo para la hipótesis alternativa H_a: $\mu < \mu_0$ sugiere una región de rechazo de la forma $z \leq c$, donde c es un número negativo adecuadamente seleccionado ($\bar{x}$ aparece muy debajo de μ_0 si y sólo si z es bastante negativo). Como Z tiene una distribución normal estándar cuando H_0 es verdadera, con $c = -z_\alpha$ da $P(\text{error de tipo I}) = \alpha$. Ésta es una prueba de *cola inferior*. Por ejemplo, $z_{.10} = 1.28$ implica que la región de rechazo $z \leq -1.28$ especifica una prueba con nivel de significancia de .10.

Por último, cuando la hipótesis alternativa es H_a: $\mu \neq \mu_0$, H_0 deberá ser rechazada si $\bar{x}$ está muy lejos de alguno de los lados de μ_0. Esto equivale a rechazar H_0 si $z \geq c$ o si $z \leq -c$. Supóngase que se desea $\alpha = .05$. Entonces,

$$\begin{aligned}.05 &= P(Z \geq c \text{ o } Z \leq -c \text{ cuando } Z \text{ tiene una distribución normal estándar})\\ &= \Phi(-c) + 1 - \Phi(c) = 2[1 - \Phi(c)]\end{aligned}$$

Por consiguiente c es tal que $1 - \Phi(c)$, el área bajo la curva z a la derecha de c, es .025 (¡y no .05!) De acuerdo con la sección 4.3 o la tabla A.3 del apéndice, $c = 1.96$ y la región de rechazo es $z \geq 1.96$ *o* $z \leq -1.96$. Con cualquier α, la región de rechazo de *dos colas* $z \geq z_{\alpha/2}$ *o* $z \leq -z_{\alpha/2}$ tiene una probabilidad α de error de tipo I (puesto que el área $\alpha/2$ está capturada debajo de cada una de las dos colas de la curva z). De nueva cuenta, la razón clave para utilizar el estadístico de prueba estandarizado Z es que como Z tiene una distribución conocida cuando H_0 es verdadera (normal estándar), es fácil de obtener una región de rechazo con probabilidad de error de tipo I deseada mediante un valor crítico apropiado.

El procedimiento de prueba en el caso I se resume en el cuadro adjunto y las regiones de rechazo correspondientes se ilustran en la figura 8.2.

Hipótesis nula: H_0: $\mu = \mu_0$

Valor del estadístico de prueba: $z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

Hipótesis alternativa	Región de rechazo para la prueba de nivel α
H_a: $\mu > \mu_0$	$z \geq z_\alpha$ (prueba de cola superior)
H_a: $\mu < \mu_0$	$z \leq -z_\alpha$ (prueba de cola inferior)
H_a: $\mu \neq \mu_0$	$z \geq z_{\alpha/2}$ o $z \leq -z_{\alpha/2}$ (prueba de dos colas)

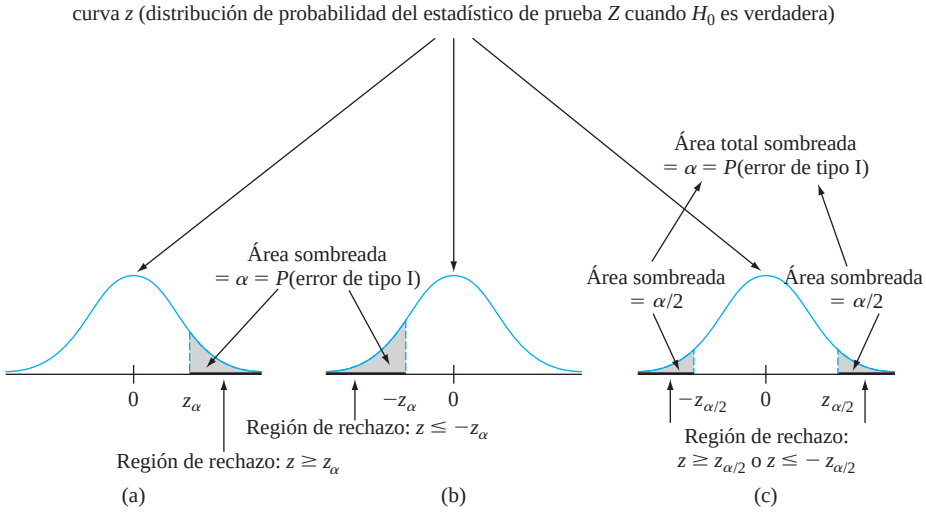


Figura 8.2 Regiones de rechazo para pruebas *z*: (a) prueba de cola superior; (b) prueba de cola inferior; (c) prueba de dos colas

Se recomienda utilizar la siguiente secuencia de pasos cuando se prueben hipótesis con respecto a un parámetro.

1. Identificar el parámetro de interés y describirlo en el contexto de la situación del problema.
2. Determinar el valor nulo y formular la hipótesis nula.
3. Formular la hipótesis alternativa apropiada
4. Dar la fórmula para el valor calculado del estadístico de prueba (sustituyendo el valor nulo y los valores conocidos de cualesquiera otros parámetros, pero *no* aquellos de cualesquiera cantidades basadas en una muestra).
5. Establecer la región de rechazo para el nivel de significancia seleccionado α.
6. Calcular cualquier cantidad muestral necesaria, sustituir en la fórmula para el valor estadístico de prueba y calcular dicho valor.
7. Decidir si H_0 debe ser rechazada y expresar esta conclusión en el contexto del problema.

La formulación de hipótesis (pasos 2 y 3) deberá ser realizada antes de examinar los datos.

Ejemplo 8.6 Un fabricante de sistemas rociadores utilizados como protección contra incendios en edificios de oficinas afirma que la temperatura de activación del sistema promedio verdadera es de 130°. Una muestra de $n = 9$ sistemas, cuando se somete a prueba, da una temperatura de activación promedio muestral de 131.08°F. Si la distribución de los tiempos de activación es normal con desviación estándar de 1.5°F, ¿contradicen los datos la afirmación del fabricante a un nivel de significancia $\alpha = .01$?

1. Parámetro de interés: μ = temperatura de activación promedio verdadera.
2. Hipótesis nula: H_0: $\mu = 130$ (valor nulo = $\mu_0 = 130$).
3. Hipótesis alternativa: H_a: $\mu \neq 130$ (un alejamiento del valor declarado en *una u otra* dirección es de interés).
4. Valor del estadístico de prueba:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{x} - 130}{1.5/\sqrt{n}}$$

5. Región de rechazo: la forma de H_a implica el uso de una prueba de dos colas con región de rechazo de $z \geq z_{.005}$ *o* $z \leq -z_{.005}$. De acuerdo con la sección 4.3 o la tabla A.3 del apéndice, $z_{.005} = 2.58$, así que se rechazaría H_0 si $z \geq 2.58$ o $z \leq -2.58$.

6. Sustituyendo $n = 9$ y $\bar{x} = 131.08$,

$$z = \frac{131.08 - 130}{1.5/\sqrt{9}} = \frac{1.08}{.5} = 2.16$$

Es decir, la media muestral observada es de un poco más de 2 desviaciones estándar sobre el valor que era de esperarse si H_0 fuera verdadera.

7. El valor calculado $z = 2.16$ no queda en la región de rechazo ($-2.58 < 2.16 < 2.58$), así que H_0 no puede ser rechazada al nivel de significancia de .01. Los datos no apoyan fuertemente la afirmación de que el promedio verdadero difiere del valor de diseño de 130. ■

***β* y determinación del tamaño de la muestra** Las pruebas z para el caso 1 se encuentran entre las pocas en estadística para las cuales existen fórmulas simples disponibles para β, la probabilidad de un error de tipo II. Considérese en primer lugar la prueba de cola superior con región de rechazo $z \geq z_\alpha$. Ésta equivale a $\bar{x} \geq \mu_0 + z_\alpha \cdot \sigma/\sqrt{n}$, por lo que H_0 no será rechazada si $\bar{x} < \mu_0 + z_\alpha \cdot \sigma/\sqrt{n}$. Ahora μ' denota un valor particular de μ que excede el valor nulo μ_0. Entonces,

$$\begin{aligned}
\beta(\mu') &= P(H_0 \text{ no es rechazada cuando } \mu = \mu') \\
&= P(\bar{X} < \mu_0 + z_\alpha \cdot \sigma/\sqrt{n} \text{ cuando } \mu = \mu') \\
&= P\left(\frac{\bar{X} - \mu'}{\sigma/\sqrt{n}} < z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}} \text{ cuando } \mu = \mu'\right) \\
&= \Phi\left(z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)
\end{aligned}$$

Conforme μ' se incrementa $\mu_0 - \mu'$ se vuelve más negativa, de modo que $\beta(\mu')$ será pequeña cuando μ' excede μ_0 en gran medida (porque el valor con el que se evalúa Φ será entonces bastante negativo). Las probabilidades de error para las pruebas de cola inferior y de dos colas se deducen de manera análoga.

Si σ es grande, la probabilidad de un error de tipo II puede ser grande con un valor alternativo de μ' que sea de interés particular para un investigador. Supóngase que se fija α y que también se especifica β para semejante valor alternativo. En el ejemplo de los sistemas rociadores, los oficiales de la compañía podrían considerar $\mu' = 132$ como un alejamiento muy sustancial de H_0: $\mu = 130$ y desear por consiguiente $\beta(132) = .10$ además de $\alpha = .01$. Más generalmente, considérense las dos restricciones $P(\text{error de tipo I}) = \alpha$ y $\beta(\mu') = \beta$ para α, μ' y β especificadas. Entonces para una prueba de cola superior, el tamaño de la muestra n debe ser elegido para satisfacer

$$\Phi\left(z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right) = \beta$$

Esto implica que

$$-z_\beta = \frac{\text{valor } z \text{ crítico que captura}}{\text{el área de cola inferior } \beta} = z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}$$

Es fácil resolver esta ecuación para la n deseada. Un argumento paralelo da el tamaño de muestra necesario para las pruebas de cola inferior y de dos colas como se resume en el siguiente recuadro.

Hipótesis alternativa	Probabilidad de error de tipo II $\beta(\mu')$ para una prueba de nivel α
H_a: $\mu > \mu_0$	$\Phi\left(z_\alpha + \dfrac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)$
H_a: $\mu < \mu_0$	$1 - \Phi\left(-z_\alpha + \dfrac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)$
H_a: $\mu \neq \mu_0$	$\Phi\left(z_{\alpha/2} + \dfrac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right) - \Phi\left(-z_{\alpha/2} + \dfrac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)$

donde $\Phi(z)$ = función de distribución acumulativa normal estándar.

El tamaño de la muestra n con el cual una prueba de nivel α también tiene $\beta(\mu') = \beta$ con el valor alternativo μ' es

$$n = \begin{cases} \left[\dfrac{\sigma(z_\alpha + z_\beta)}{\mu_0 - \mu'}\right]^2 & \text{prueba de una cola (superior o inferior)} \\ \left[\dfrac{\sigma(z_{\alpha/2} + z_\beta)}{\mu_0 - \mu'}\right]^2 & \text{para una prueba de dos colas (una solución aproximada)} \end{cases}$$

Ejemplo 8.7 Sea μ la vida promedio verdadera de la banda de rodamiento de un cierto tipo de neumático. Considere poner a prueba H_0: $\mu = 30{,}000$ contra H_a: $\mu > 30{,}000$ basado en un tamaño de muestra $n = 16$ de una distribución de población normal con $\sigma = 1500$. Una prueba con $\alpha = .01$ requiere $z_\alpha = z_{.01} = 2.33$. La probabilidad de cometer un error de tipo II cuando $\mu = 31{,}000$ es

$$\beta(31{,}000) = \Phi\left(2.33 + \frac{30{,}000 - 31{,}000}{1500/\sqrt{16}}\right) = \Phi(-.34) = .3669$$

Como $z_{.1} = 1.28$, el requerimiento de que el nivel de prueba .01 también tenga $\beta(31{,}000) = .1$ necesita

$$n = \left[\frac{1500(2.33 + 1.28)}{30{,}000 - 31{,}000}\right]^2 = (-5.42)^2 = 29.32$$

El tamaño de muestra debe ser un entero, por lo tanto se deberán utilizar 30 neumáticos. ■

Caso II: Pruebas con muestras grandes

Cuando el tamaño de muestra es grande, las pruebas z en el caso I son fáciles de modificar para dar procedimientos de prueba válidos sin requerir una distribución de población normal o σ conocida. El resultado clave se utilizó en el capítulo 7 para justificar intervalos de confianza para muestra grande: una n grande implica que la variable estandarizada

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

tiene *aproximadamente* una distribución normal estándar. La sustitución del valor nulo μ_0 en lugar de μ da el estadístico de prueba

$$Z = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

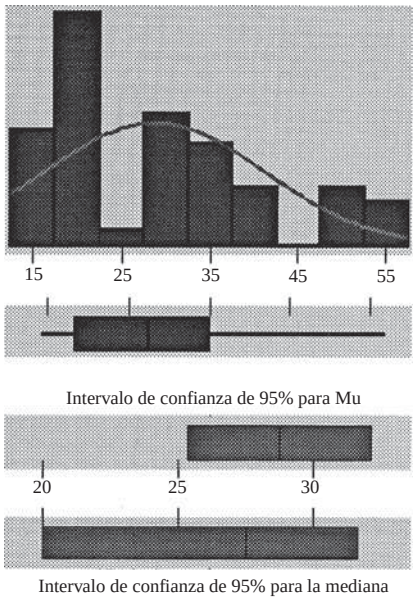
que tiene de manera aproximada una distribución normal estándar cuando H_0 es verdadera. El uso de las regiones de rechazo dadas previamente para el caso I (p. ej., $z \geq z_\alpha$ cuando la hipótesis alternativa es H_a: $\mu > \mu_0$) resulta entonces en procedimientos de prueba con

los cuales el nivel de significancia es casi (y no exactamente) α. Se utilizará de nuevo la regla empírica $n > 40$ para caracterizar un tamaño de muestra grande.

Ejemplo 8.8 Se utiliza un penetrómetro cónico dinámico para medir la resistencia de un material a la penetración (mm/golpe), a medida que el cono es insertado en pavimento o subsuelo. Suponga que para una aplicación particular, se requiere que el valor de penetración cónica dinámica promedio verdadera para un cierto tipo de pavimento sea menor que 30. El pavimento no será utilizado a menos que exista evidencia concluyente de que la especificación se satisfizo. Formule y pruebe las hipótesis apropiadas utilizando los datos siguientes ("Probabilistic Model for the Analysis of Dynamic Cone Penetrometer Test Values in Pavement Structure Evaluation", *J. of Testing and Evaluation*, 1999: 7–14:

14.1	14.5	15.5	16.0	16.0	16.7	16.9	17.1	17.5	17.8
17.8	18.1	18.2	18.3	18.3	19.0	19.2	19.4	20.0	20.0
20.8	20.8	21.0	21.5	23.5	27.5	27.5	28.0	28.3	30.0
30.0	31.6	31.7	31.7	32.5	33.5	33.9	35.0	35.0	35.0
36.7	40.0	40.0	41.3	41.7	47.5	50.0	51.0	51.8	54.4
55.0	57.0								

La figura 8.3 muestra un resumen descriptivo obtenido con Minitab. La penetración cónica dinámica media muestral es menor que 30. Sin embargo, existe una cantidad sustancial de variación en los datos (coeficiente de variación muestral $= s/\bar{x} = .4265$), de modo que el hecho de que la media sea menor que el valor de corte de la especificación de diseño puede ser simplemente una consecuencia de la variabilidad muestral. Obsérvese que el histograma no se asemeja en absoluto a una curva normal (y una gráfica de probabilidad normal no exhibe un patrón lineal), aunque las pruebas z con muestras grandes no requieren una distribución de población normal.



Estadística descriptiva

Variable: penetración cónica dinámica

Prueba de normalidad de Anderson-Darling

A al cuadrado	1.902
Valor P:	0.000
Media	28.7615
Desv. Estd.	12.2647
Varianza	150.423
Asimetría	0.808264
Kurtosis	–3.9E–01
N	52
Mínimo	14.1000
1er. cuartil	18.2250
Mediana	27.5000
3er. cuartil	35.0000
Máximo	57.0000

Intervalo de confianza de 95% para Mu

25.3470	3.21761

Intervalo de confianza de 95% para sigma

10.2784	15.2098

Intervalo de confianza de 95% para la mediana

20.0000	31.7000

Figura 8.3 Resumen descriptivo generado por Minitab para los datos de penetración cónica dinámica del ejemplo 8.8

1. μ = valor de penetración cónica dinámica promedio verdadero
2. H_0: $\mu = 30$

3. H_a: $\mu < 30$ (por consiguiente el pavimento no será utilizado a menos que la hipótesis nula sea rechazada)
4. $z = \dfrac{\bar{x} - 30}{s/\sqrt{n}}$
5. Una prueba con nivel de significancia de .05 rechaza a H_0 cuando $z \leq -1.645$ (una prueba de cola inferior).
6. Con $n = 52$, $\bar{x} = 28.76$ y $s = 12.2647$,

$$z = \frac{28.76 - 30}{12.2647/\sqrt{52}} = \frac{-1.24}{1.701} = -.73$$

7. Como $-.73 > -1.645$, H_0 no puede ser rechazada. No se cuenta con evidencia precisa para concluir que $\mu < 30$; el uso del pavimento no se justifica. ■

La determinación de β y el tamaño de muestra necesario para estas pruebas con muestra grande pueden basarse en la especificación de un valor adecuado de σ y en el uso de las fórmulas para el caso I (aun cuando se utilice s en la prueba) o en el uso de la metodología que se introducirá en breve en conexión con el caso III.

Caso III: Una distribución de población normal

Cuando n es pequeño, el teorema del límite central TLC ya no puede ser invocado para justificar el uso de una prueba con muestra grande. Esta misma dificultad se presentó al obtener un intervalo de confianza (IC) con muestra pequeña para μ en el capítulo 7. El método utilizado aquí es el mismo que se empleó allí: se supondrá que la distribución de población es por lo menos aproximadamente normal y se describirán los procedimientos de prueba cuya validez se fundamenta en esta suposición. Si un investigador tiene una buena razón para creer que la distribución de población es bastante no normal, se puede utilizar una prueba libre de distribución del capítulo 15. Alternativamente, un estadístico puede ser consultado en cuanto a procedimientos válidos para familias específicas de distribuciones de población aparte de la familia normal. O se puede desarrollar un procedimiento bootstrap.

En el capítulo 7 se utilizó el resultado clave en el cual están basadas las pruebas con una media de población normal para obtener el intervalo de confianza t para una muestra: si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria de una distribución normal, la variable estandarizada

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

tiene una distribución t con $n - 1$ grados de libertad (gl). Considérese poner a prueba H_0: $\mu = \mu_0$ contra H_a: $\mu > \mu_0$ utilizando el estadístico de prueba $T = (\bar{X} - \mu_0)/(S/\sqrt{n})$. Es decir, el estadístico de prueba resulta de estandarizar $\bar{X}$ conforme a la suposición de que H_0 es verdadera (utilizando $S/\sqrt{n}$, la desviación estándar estimada de $\bar{X}$, en lugar de $\sigma/\sqrt{n}$). Cuando H_0 es verdadera, el estadístico de prueba tiene una distribución t con $n - 1$ grados de libertad. El conocimiento de la distribución del estadístico de prueba cuando H_0 es verdadera (la "distribución nula") permite construir una región de rechazo para la cual la probabilidad de error de tipo I se controla al nivel deseado. En particular, el uso del valor crítico t de cola superior $t_{\alpha,n-1}$ para especificar la región de rechazo $t \geq t_{\alpha,n-1}$ implica que

$$\begin{aligned} P(\text{error de tipo I}) &= P(H_0 \text{ es rechazada cuando es verdadera}) \\ &= P(T \geq t_{\alpha,n-1} \text{ cuando } T \text{ tiene una distribución } t \text{ con } n-1 \\ &\quad \text{grados de libertad}) \\ &= \alpha \end{aligned}$$

El estadístico de prueba es en realidad el mismo del caso de muestra grande pero se designa T para recalcar que su distribución nula es una distribución t con $n - 1$ grados de

libertad en lugar de la distribución normal estándar (z). La región de rechazo para la prueba *t* difiere de aquella para la prueba *z* sólo en que un valor crítico $t_{\alpha,n-1}$ reemplaza al valor crítico z_α de *z*. Comentarios similares se aplican a alternativas para las cuales una prueba de cola inferior o de dos colas es apropiada.

Prueba *t* con una muestra

Hipótesis nula: H_0: $\mu = \mu_0$

Valor estadístico de prueba: $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$

Hipótesis alternativa	Región de rechazo para una prueba de nivel α
H_a: $\mu > \mu_0$	$t \geq t_{\alpha,n-1}$ (cola superior)
H_a: $\mu < \mu_0$	$t \leq -t_{\alpha,n-1}$ (cola inferior)
H_a: $\mu \neq \mu_0$	$t \geq t_{\alpha/2,n-1}$ o $t \leq -t_{\alpha/2,n-1}$ (dos colas)

Ejemplo 8.9 El glicerol es un importante subproducto de la fermentación de etanol en la producción de vino y contribuye a la dulzura, cuerpo y la plenitud de los vinos. El artículo "A Rapid and Simple Method for Simultaneous Determination of Glycerol, Fructose, and Glucose in Wine" (*American J. of Enology and Viticulture*, 2007: 279−283) incluye las siguientes observaciones sobre la concentración de glicerol (mg/ml) para muestras de vinos blancos de calidad estándar (sin certificar): 2.67, 4.62, 4.14, 3.81, 3.83. Supongamos que el valor de la concentración deseada es de 4. ¿Los datos de la muestra indican que la concentración promedio real es algo más que el valor deseado? El siguiente diagrama de probabilidad normal de Minitab proporciona un fuerte apoyo para suponer que la distribución de la población de la concentración de glicerol es normal. Se llevará a cabo una prueba de hipótesis adecuada utilizando la prueba *t* de una muestra con un nivel de significancia de .05.

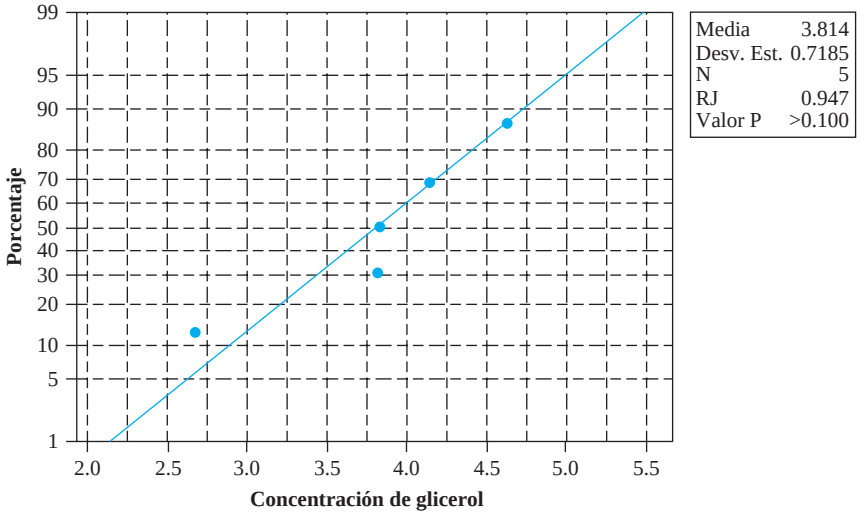


Figura 8.4 Gráfica de probabilidad normal para los datos del ejemplo 8.9

1. $\mu =$ promedio real de la concentración de glicerol
2. H_0: $\mu = 4$
3. H_a: $\mu \neq 4$

4. $t = \dfrac{\bar{x} - 4}{s/\sqrt{n}}$

5. La desigualdad en H_a implica que una prueba de dos colas es apropiada, lo que requiere $t_{\alpha/2,n-1} = t_{025,4} = 2.776$. Por lo tanto H_0 se rechaza si $t \geq 2.776$ o $t \leq -2.776$.

6. $\sum x_i = 19.07$ y $\sum x_i^2 = 74.7979$ con las cuales $\bar{x} = 3.814$, $s = .718$ y el error estándar estimado de la media es $s/\sqrt{n} = .321$. El valor del estadístico de prueba es entonces $t = (3.814 - 4)/.321 = -.58$.

7. Es evidente que $t = -.58$ no se encuentra en la región de rechazo para un nivel de significancia de .05. Todavía es posible que $\mu = 4$. La desviación de la media de la muestra 3.814 de su valor esperado 4 cuando H_0 es cierta se puede atribuir simplemente a la variabilidad del muestreo en lugar de que H_0 sea falsa.

Los datos de salida de Minitab resultan de una solicitud para realizar una prueba *t* de dos colas de una muestra y presenta valores calculados idénticos a los que se acaban de obtener. El hecho de que el último número en la salida, el "valor P" sea superior a .05 (y cualquier otro nivel de significancia razonable) implica que la hipótesis nula no puede rechazarse. Esto se analiza con detalle en la sección 8.4.

```
Test of mu = 4 vs not = 4
Variable    N   Mean  StDev  SE Mean  95% CI             T      P
glyc conc   5  3.814  0.718    0.321  (2.922, 4.706)  −0.58  0.594
```

■

***β* y determinación del tamaño de la muestra** El cálculo de β con el valor alternativo μ' en el caso I se realizó expresando la región de rechazo en función de $\bar{x}$ (p. ej., $\bar{x} \geq \mu_0 + z_\alpha \cdot \sigma/\sqrt{n}$) y luego restando μ' para estandarizar correctamente. Un método equivalente implica observar que cuando $\mu = \mu'$, el estadístico de prueba $Z = (\bar{X} - \mu_0)/(\sigma/\sqrt{n})$ sigue teniendo una distribución normal con varianza 1, pero ahora el valor medio de Z está dado por $(\mu' - \mu_0)/(\sigma/\sqrt{n})$. Es decir, cuando $\mu = \mu'$, el estadístico de prueba sigue teniendo una distribución normal pero no la distribución normal estándar. Por eso, $\beta(\mu')$ es un área bajo la curva normal correspondiente al valor medio $(\mu' - \mu_0)/(\sigma/\sqrt{n})$ y varianza 1. Tanto α como β implican trabajar con variables normalmente distribuidas.

El cálculo de $\beta(\mu')$ para la prueba *t* es mucho menos directo. Esto es porque la distribución del estadístico de prueba $T = (\bar{X} - \mu_0)/(S/\sqrt{n})$ es bastante complicada cuando H_0 es falsa y H_a es verdadera. Por consiguiente, en una prueba de cola superior, determinar

$$\beta(\mu') = P(T < t_{\alpha,n-1} \text{ cuando } \mu = \mu' \text{ en lugar de } \mu_0)$$

implica integrar una desagradable función de densidad. Esto debe hacerse numéricamente. Los resultados se resumen en gráficas de β que aparecen en la tabla A.17 del apéndice. Existen cuatro juegos de gráficas, correspondientes a pruebas de una cola a nivel .05 y nivel .01 y pruebas de dos colas a los mismos niveles.

Para entender cómo se utilizan estas gráficas, obsérvese primero que tanto β como el tamaño de muestra necesario n en el caso I son funciones no sólo de la diferencia absoluta $|\mu_0 - \mu'|$ sino de $d = |\mu_0 - \mu'|/\sigma$. Supóngase, por ejemplo, que $|\mu_0 - \mu'| = 10$. Este alejamiento de H_0 será mucho más fácil de descubrir (β más pequeña) cuando $\sigma = 2$, en cuyo caso μ_0 y μ' están a 5 desviaciones estándar de la población una de otra, que cuando $\sigma = 10$. El hecho de que β para la prueba *t* dependa de d y no sólo de $|\mu_0 - \mu'|$ es desafortunado, puesto que para utilizar las gráficas se debe tener alguna idea del valor verdadero de σ. Una suposición conservadora (grande) para σ dará por resultado un valor conservador (grande) de $\beta(\mu')$ y una estimación conservadora del tamaño de muestra necesario para α y $\beta(\mu')$ prescritas.

Una vez que se seleccionan la μ' alternativa y el valor de σ, se calcula d y su valor se localiza sobre el eje horizontal del conjunto de curvas pertinente. El valor de β es la altura de la curva con $n - 1$ grados de libertad por encima del valor de d (es necesaria una interpolación visual si $n - 1$ no es un valor para el cual la curva correspondiente aparece), como se ilustra en la figura 8.5.

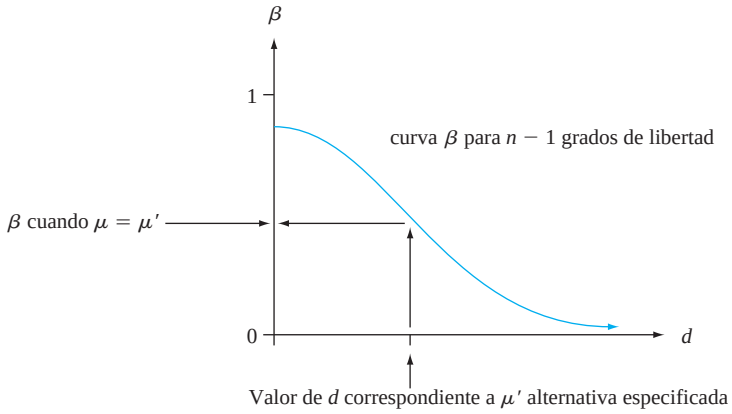


Figura 8.5 Curva β típica de la prueba t

En lugar de fijar n (es decir, $n - 1$) y por consiguiente la curva particular en donde se lee β) se podría prescribir tanto α (.05 o .01 en este caso) y un valor de β para las μ' y σ seleccionadas. Después de calcular d, se localiza el punto (d, β) en el conjunto de gráficas pertinentes. La curva debajo y más próxima a este punto da $n - 1$ y por consiguiente n (de nuevo con frecuencia se requiere interpolación).

Ejemplo 8.10 Se supone que la caída de voltaje promedio verdadera entre el colector y el emisor de transistores bipolares de compuerta aislados de cierto tipo es cuando mucho de 2.5 volts. Un investigador selecciona una muestra de $n = 10$ de esos transistores y utiliza los voltajes resultantes para probar H_0: $\mu = 2.5$ contra H_a: $\mu > 2.5$ por medio de una prueba t con nivel de significancia $\alpha = .05$. Si la desviación estándar de la distribución de voltaje es $\sigma = .100$, ¿qué tan probable es que H_0 no será rechazada cuando en realidad $\mu = 2.6$? Con $d = |2.5 - 2.6|/.100 = 1.0$, el punto sobre la curva β con 9 grados de libertad para una prueba de una cola con $\alpha = .05$ por encima de 1.0 tiene aproximadamente .1 de altura, por lo tanto $\beta \approx .1$. El investigador podría pensar que éste es un valor de β demasiado grande con semejante alejamiento sustancial de H_0 y puede que desee tener $\beta = .05$ con este valor alternativo de μ. Como $d = 1.0$, el punto $(d, \beta) = (1.0, .05)$ debe ser localizado. Este punto se aproxima mucho a la curva de 14 grados de libertad, por lo tanto con $n = 15$ se obtendrá tanto $\alpha = .05$ como $\beta = .05$ cuando el valor de μ es 2.6 y $\sigma = .10$. Un valor más grande de σ daría una β más grande para esta alternativa y un valor alternativo de μ más cercano a 2.5 también daría por resultado un valor incrementado de β. ■

La mayoría de los programas de computadora estadísticos también calcularán probabilidades de error de tipo II. Por lo general, trabajan en términos de **potencia**, que es simplemente $1 - \beta$. Un valor pequeño de β (cerca de 0) es equivalente a una potencia grande (cerca de 1). Una prueba de *gran alcance* es la que tiene gran potencia y por tanto buena capacidad para detectar cuándo la hipótesis nula es falsa.

Como ejemplo, se le pide a Minitab que determine la potencia de la prueba de cola superior del ejemplo 8.10 para tres tamaños de muestra de 5, 10 y 15 cuando $\alpha = .05$, $\sigma = .10$ y el valor de μ es en realidad 2.6 en vez del valor nulo de 2.5, una "diferencia" de

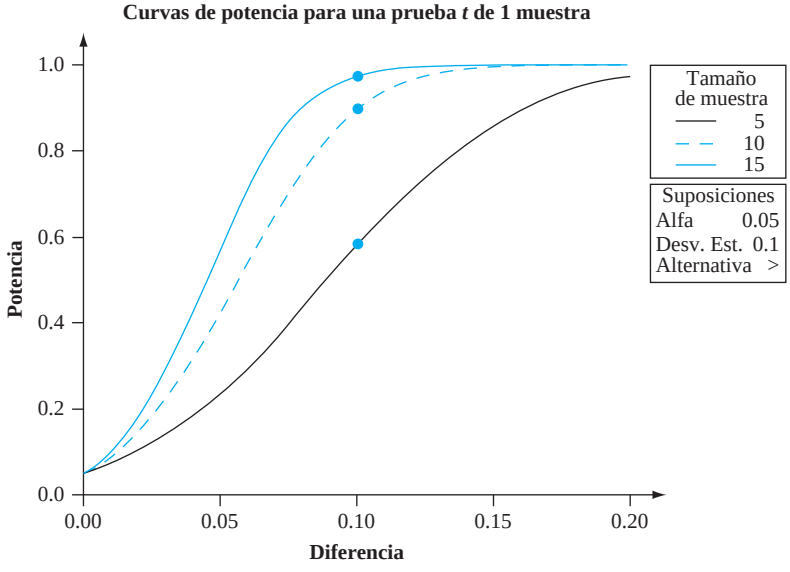
$2.6 - 2.5 = .1$. También se le pidió al software determinar el tamaño de la muestra necesario para una potencia de .9 ($\beta = .1$) y .95. Los datos de salida se dan a continuación.

```
Power and Sample Size
Testing mean = null (versus > null)
Calculating power for mean = null + difference
Alpha = 0.05 Assumed standard deviation = 0.1

            Sample
Difference     Size    Power
0.1             5      0.579737
0.1            10      0.897517
0.1            15      0.978916

               Sample Target
                                    Actual
Difference     Size    Power        Power
0.1            11      0.90         0.924489
0.1            13      0.95         0.959703
```

La potencia para el tamaño de muestra $n = 10$ es un poco menor que .9. Así que si se insiste en que la potencia sea de por lo menos .9, es necesaria una muestra de tamaño 11 y la potencia real para n es aproximadamente .92. El software dice que para una potencia objetivo de .95, es necesario un tamaño de muestra de $n = 13$, mientras que echando un vistazo a nuestras curvas β dio 15. Cuando está disponible, este tipo de software es más confiable que las curvas. Por último, Minitab ahora también proporciona curvas de potencia para los tamaños de muestra determinados, tal como se muestra en la figura 8.6. Estas curvas muestran cómo aumenta la potencia para cada tamaño de muestra a medida que el valor real de μ se desplaza más allá y más lejos del valor nulo.



EJERCICIOS Sección 8.2 (15–36)

15. Sea un estadístico de prueba Z que tiene una distribución normal estándar cuando H_0 es verdadera. Dé el nivel de significancia en cada una de las siguientes situaciones:

a. H_a: $\mu > \mu_0$, región de rechazo $z \geq 1.88$
b. H_a: $\mu < \mu_0$, región de rechazo $z \leq -2.75$
c. H_a: $\mu \neq \mu_0$, región de rechazo $z \geq 2.88$ *o* $z \leq -2.88$

16. Sea el estadístico de prueba T que tiene una distribución t cuando H_0 es verdadera. Dé el nivel de significancia en cada una de las situaciones:

a. H_a: $\mu > \mu_0$, grados de libertad $= 15$, región de rechazo $t \geq 3.733$

b. H_a: $\mu < \mu_0$, $n = 24$, región de rechazo $t \leq -2.500$

c. H_a: $\mu \neq \mu_0$, $n = 31$, región de rechazo $t \geq 1.697$ o $t \leq -1.697$

17. Responda las siguientes preguntas en relación con el problema de los neumáticos en el ejemplo 8.7.

a. Si $\bar{x} = 30{,}960$ y se utiliza una prueba de nivel $\alpha = .01$, ¿cuál es la decisión?

b. Si utiliza una prueba de nivel .01, ¿cuál es $\beta(30{,}500)$?

c. Si se utiliza una prueba de nivel .01 y también se requiere que $\beta(30{,}500) = .05$, ¿qué tamaño de muestra n es necesario?

d. Si $\bar{x} = 30{,}960$ ¿cuál es la α más pequeña con la cual H_0 puede ser rechazada (con base en $n = 16$)?

18. Reconsidere la situación de secado de pintura del ejemplo 8.2, en el cual el tiempo de secado para un espécimen de prueba está normalmente distribuido con $\sigma = 9$. Las hipótesis H_0: $\mu = 75$ contra H_a: $\mu < 75$ tienen que ser probadas con una muestra aleatoria de $n = 25$ observaciones.

a. ¿A cuántas desviaciones estándar (de $\bar{X}$) por debajo del valor nulo se encuentra $\bar{x} = 72.3$?

b. Si $\bar{x} = 72.3$, ¿cuál es la conclusión si utiliza $\alpha = .01$?

c. ¿Cuál es α para el procedimiento de prueba que rechaza H_0 cuando $z \leq -2.88$?

d. Con el procedimiento de prueba del inciso (c), ¿cuál es $\beta(70)$?

e. Si se utiliza el procedimiento de prueba del inciso (c), ¿qué n es necesaria para garantizar $\beta(70) = .01$?

f. Si se utiliza un prueba de nivel .01 con $n = 100$, ¿cuál es la probabilidad de un error de tipo I cuando $\mu = 76$?

19. Se determinó el punto de fusión de cada una de las 16 muestras de una marca de aceite vegetal hidrogenado y el resultado fue $\bar{x} = 94.32$. Suponiendo que la distribución del punto de fusión es normal con $\sigma = 1.20$.

a. Probar H_0: $\mu = 95$ contra H_a: $\mu \neq 95$ por medio de una prueba de dos colas de nivel .01.

b. Si se utiliza una prueba de nivel .01, ¿cuál es $\beta(94)$, la probabilidad de un error de tipo II cuando $\mu = 94$?

c. ¿Qué valor de n es necesario para garantizar que $\beta(94) = .1$ cuando $\alpha = .01$?

20. Se anuncia que focos de un tipo duran un promedio de 750 horas. El precio de estos focos es muy favorable por lo que un cliente potencial ha decidido continuar con un convenio de compra hasta que concluyentemente se demuestre que la duración promedio verdadera sea menor que la anunciada. Se seleccionó una muestra aleatoria de 50 focos, se determinó la duración de cada uno, se probaron las hipótesis apropiadas con Minitab y se obtuvieron los siguientes resultados.

```
Variable   N    Mean  StDev  SEMean      Z  P-Value
lifetime  50  738.44  38.20    5.40  -2.14    0.016
```

¿Qué conclusión sería apropiada para un nivel de significancia de .05? ¿Un nivel de significancia de .01? ¿Qué nivel de significancia y conclusión recomendaría?

21. Se supone que el diámetro promedio verdadero de cojinetes de bolas de un tipo es de .5 pulg. Se realizará una prueba t con una muestra para ver si éste es el caso. ¿Qué conclusión es apropiada en cada una de las siguientes situaciones?

a. $n = 13$, $t = 1.6$, $\alpha = .05$

b. $n = 13$, $t = -1.6$, $\alpha = .05$

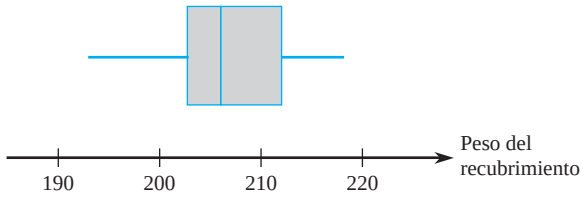
c. $n = 25$, $t = -2.6$, $\alpha = .01$

d. $n = 25$, $t = -3.9$

22. El artículo "The Foreman's View of Quality Control" (*Quality Engr.*, 1990: 257–280) describe una investigación de pesos de recubrimiento de grandes tuberías resultantes de un proceso de galvanizado. Los estándares de producción demandan un peso promedio verdadero de 200 lb por tubería. El resumen y la gráfica de caja descriptivos adjuntos fueron producidos por Minitab.

```
Variable N   Mean  Median TrMean StDev  SEMean
ctg wt   30 206.73 206.00 206.81  6.35    1.16

Variable    Min     Max     Q1      Q3
ctg wt      193.00  218.00  202.75  212.00
```



a. ¿Qué sugiere la gráfica de caja sobre el estado de la especificación de peso de recubrimiento promedio verdadero?

b. Una gráfica de probabilidad normal de los datos resultó bastante recta. Use los datos de salida descriptivos para probar las hipótesis apropiadas.

23. El ejercicio 36 del capítulo 1 dio $n = 26$ observaciones sobre el tiempo de escape (s) de trabajadores petroleros en un ejercicio simulado, con media y desviación estándar muestrales de 370.69 y 24.36, respectivamente. Suponga que los investigadores creyeron de antemano que el tiempo de escape promedio verdadero sería cuando mucho de 6 min. ¿Contradicen los datos esta creencia anticipada? Suponiendo normalidad, pruebe las hipótesis apropiadas con un nivel de significancia de .05.

24. Reconsidere las observaciones muestrales sobre viscosidad estabilizada de especímenes de asfalto introducidos en el ejercicio 46 del capítulo 1 (2781, 2900, 3013, 2856 y 2888). Suponga que para una aplicación particular se requiere que la viscosidad promedio verdadera sea de 3000. ¿Parece haber sido satisfecho este requerimiento? Formule y pruebe las hipótesis apropiadas.

25. El porcentaje deseado de SiO_2 en cierto tipo de cemento aluminoso es de 5.5. Para comprobar si el porcentaje promedio verdadero es de 5.5 en una instalación de producción particular, se analizaron 16 muestras obtenidas de manera independiente. Suponga que el porcentaje de SiO_2 en una muestra está normalmente distribuido con $\sigma = .3$ y que $\bar{x} = 5.25$.

a. ¿Indica esto concluyentemente que el porcentaje promedio verdadero difiere de 5.5? Realice el análisis siguiendo la secuencia de pasos sugerida en el texto.

b. Si el porcentaje promedio verdadero es $\mu = 5.6$ y se utiliza una prueba de nivel $\alpha = .01$ con $n = 16$, ¿cuál es la probabilidad de descubrir este alejamiento de H_0?

c. ¿Qué valor de n se requiere para satisfacer $\alpha = .01$ y $\beta(5.6) = .01$?

26. Para obtener información sobre las propiedades de resistencia a la corrosión de un tipo de conducto de acero, se enterraron 45 especímenes en el suelo durante 2 años. Se midió entonces la penetración máxima (en mils) en cada espécimen y se obtuvo una penetración promedio muestral de $\bar{x} = 52.7$ y una desviación estándar muestral de $s = 4.8$. Los conductos se fabricaron con la especificación de que la penetración promedio verdadera sea cuando mucho de 50 mils. Se utilizarán a menos que se pueda demostrar concluyentemente que la especificación no ha sido satisfecha. ¿Qué concluiría?

27. La identificación automática de los límites de estructuras significativas en una imagen médica es un área de investigación continua. El artículo "Automatic Segmentation of Medical Images Using Image Registration: Diagnostic and Simulation Applications" (*J. of Medical Engr. and Tech.*, 2005: 53–63) discutió una nueva técnica para realizar la identificación mencionada. Una medida de la precisión de la región automática es el desplazamiento lineal promedio. El artículo dio las siguientes observaciones de desplazamiento lineal promedio con una muestra de 49 riñones (unidades de dimensiones en pixeles).

```
1.38   0.44   1.09   0.75   0.66   1.28   0.51
0.39   0.70   0.46   0.54   0.83   0.58   0.64
1.30   0.57   0.43   0.62   1.00   1.05   0.82
1.10   0.65   0.99   0.56   0.56   0.64   0.45
0.82   1.06   0.41   0.58   0.66   0.54   0.83
0.59   0.51   1.04   0.85   0.45   0.52   0.58
1.11   0.34   1.25   0.38   1.44   1.28   0.51
```

a. Resuma y describa los datos.

b. ¿Es factible que el desplazamiento lineal promedio esté por lo menos normalmente distribuido en forma aproximada? ¿Se debe suponer normalidad antes de calcular un intervalo de confianza para el desplazamiento lineal promedio verdadero o probar las hipótesis en cuanto a desplazamiento lineal promedio verdadero? Explique.

c. Los autores comentaron que en la mayoría de los casos el desplazamiento lineal promedio es del orden de 1.0 o mejor. ¿Proporcionan en realidad los datos una fuerte evidencia para concluir que el desplazamiento lineal promedio en estas circunstancias es menor que 1.0? Efectúe una prueba apropiada de hipótesis.

d. Calcule un límite de confianza superior para el desplazamiento lineal promedio verdadero utilizando un nivel de confianza de 95% e interprete este límite.

28. La cirugía menor de caballos en condiciones de campo requiere un anestésico de corta duración confiable que produzca una buena relajación muscular, cambios cardiovasculares y respiratorios mínimos y una rápida y tranquila recuperación con mínimos efectos secundarios de modo que los caballos puedan ser dejados sin atención. El artículo "A Field Trial of Ketamine Anesthesia in the Horse" (*Equine Vet. J.*, 1984: 176–179) reporta que con una muestra de $n = 73$ caballos a los cuales se les administró ketamina en ciertas condiciones, el tiempo de reclinación lateral (echado) promedio muestral fue de 18.86 min y la desviación estándar de 8.6 min. ¿Sugieren estos datos que el tiempo de reclinación lateral promedio verdadera en estas condiciones es menor que 20 min? Pruebe las hipótesis apropiadas a un nivel de significancia de .10.

29. El artículo "Uncertainty Estimation in Railway Track Life-Cycle Cost" (*J. of Rail and Rapid Transit*, 2009) presenta los siguientes datos sobre el tiempo de reparación (minutos) de la rotura de un carril alto en una vía curva del tren de cierta línea de ferrocarril.

```
159 120 480 149 270 547 340 43 228 202 240 218
```

Una gráfica de probabilidad normal de los datos muestra un patrón bastante lineal, por lo que es factible que la distribución de la población del tiempo de reparación sea por lo menos aproximadamente normal. La desviación media y estándar de la muestra son 249.7 y 145.1, respectivamente.

a. ¿Hay pruebas de peso para concluir que de verdad el tiempo medio de reparación sea superior a 200 minutos? Lleve a cabo una prueba de hipótesis con un nivel de significancia de .05.

b. Usando $\sigma = 150$, ¿cuál es la probabilidad de error tipo II de la prueba utilizada en el inciso (a) cuando el tiempo promedio de reparación verdadero es en realidad 300 minutos? Es decir, ¿cuál es $\beta(300)$?

30. ¿Alguna vez se ha visto frustrado porque no puede conseguir un contenedor de algún tipo del que se pueda liberar la última parte de su contenido? El artículo "Shake, Rattle, and Squeeze: How Much Is Left in That Container?" (*Consumer Reports*, May 2009: 8), informó sobre una investigación de este tema para varios productos de consumo. Supongamos que cinco tubos de 6.0 onzas de pasta de dientes de una marca en particular son seleccionados al azar y se les aprieta hasta que no haya más pasta de dientes que salga. Luego, se corta cada tubo y la cantidad restante se pesa, dando lugar a los siguientes datos (en consistencia con lo que el citado artículo informaba): .53, .65, .46, .50, .37. ¿Parece que la cantidad promedio restante real es inferior al 10% del contenido neto anunciado?

a. Compruebe la validez de los supuestos necesarios para probar la hipótesis apropiada.

b. Lleve a cabo una prueba de las hipótesis adecuadas con un nivel de significancia de .05. ¿Cambiaría su conclusión si se ha utilizado un nivel de significancia de .01?

c. Describa en contexto los tipos de errores I y II, y diga qué error podría haberse cometido para llegar a una conclusión.

31. Un lugar de trabajo seguro y bien diseñado puede contribuir en gran medida al aumento de la productividad. Es especialmente importante que los trabajadores no realicen tareas que excedan sus capacidades, tales como cargar. Los siguientes datos sobre el peso máximo de carga (PMC, en kg) para una frecuencia de cuatro cargas/min se presentaron en el artículo "The Effects of Speed, Frequency, and Load on Measured Hand Forces for a Floor-to-Knuckle Lifting Task" (*Ergonomics*, 1992: 833−843), los sujetos fueron seleccionados al azar de una población de hombres sanos con edades de 18 a 30 años. Suponiendo que el PMC se distribuye normalmente, ¿los datos sugieren que la media poblacional del PMC supera los 25? Lleve a cabo una prueba con un nivel de significancia de .05.

```
25.8   36.6   26.3   21.8   27.2
```

32. La cantidad diaria recomendada de zinc en la dieta entre los varones mayores de 50 años de edad es de 15 mg/día. El artículo "Nutrient Intakes and Dietary Patterns of Older Americans: A National Study" (*J. of Gerontology*, 1992: M145−150) presenta el siguiente resumen de los datos sobre el consumo de zinc en una muestra de varones con edades entre 65–74 años: $n = 115$, $\bar{x} = 11.3$, y $s = 6.43$. ¿Estos datos indican que la ingesta de zinc

diaria promedio en la población de hombres de todas las edades de 65 a 74 años cae por debajo de la cantidad recomendada?

33. Reconsidere los datos del ejemplo que muestran la proporción de gastos (%) de los fondos de crecimiento de gran capitalización mutua presentados por primera vez en el ejercicio 1.53.

```
0.52 1.06 1.26 2.17 1.55 0.99 1.10 1.07 1.81 2.05
0.91 0.79 1.39 0.62 1.52 1.02 1.10 1.78 1.01 1.15
```

Una gráfica de probabilidad normal muestra un patrón bastante lineal.

a. ¿Hay pruebas de peso para concluir que la media poblacional de la proporción de gastos excede el 1%? Lleve a cabo una prueba de las hipótesis pertinentes con un nivel de significancia de .01.

b. Volviendo al inciso (a), describa en contexto el tipo de errores I y II y diga qué error podría haberse cometido para llegar a su conclusión. La fuente de la cual se obtuvieron los datos informó que $\mu = 1.33$ para la población de todos los 762 fondos. Así que, ¿ha cometido un error al llegar a su conclusión?

c. Suponiendo que $\sigma = .5$, determine e interprete la potencia de la prueba en el inciso (a) para el valor real de μ indicado en el inciso (b).

34. Una muestra de 12 detectores de radón de un cierto tipo fue seleccionada y cada uno fue expuesto a 100 pCi/L de radón. Las lecturas resultantes son las siguientes:

105.6	90.9	91.2	96.9	96.5	91.3
100.1	105.0	99.6	107.7	103.3	92.4

a. ¿Estos datos sugieren que la lectura media de la población en estas condiciones difiere de 100? Establezca y ponga a prueba las hipótesis adecuadas con $\alpha = .05$.

b. Suponga que antes del experimento se había supuesto un valor de $\sigma = 7.5$. ¿Cuántas determinaciones habrían sido apropiadas entonces para obtener $\beta = .10$ para la $\mu = 95$ alternativa?

35. Demuestre que para cualquier $\Delta > 0$, cuando la distribución de la población es normal y se conoce σ, la prueba de dos colas satisface $\beta(\mu_0 - \Delta) = \beta(\mu_0 + \Delta)$, de modo que $\beta(\mu')$ es simétrica respecto a μ_0.

36. Para un valor μ' alternativo fijo, demuestre que $\beta(\mu') \to 0$ cuando $n \to \infty$, para una prueba z de una cola o de dos colas en el caso de una distribución normal de la población con σ conocida.

8.3 Pruebas relacionadas con una proporción de población

Sea p la proporción de individuos u objetos en una población que poseen una propiedad especial (p. ej., carros con transmisión manual o fumadores que fuman cigarrillos con filtro). Si un individuo u objeto con la propiedad es etiquetado como éxito (S), entonces p es la proporción de éxitos de la población. Las pruebas relacionadas con p se basarán en una muestra aleatoria de tamaño n de la población. Siempre que n sea pequeña con respecto al tamaño de la población, X (el número de éxitos en la muestra) tiene (aproximadamente) una distribución binomial. Además, si n es grande [$np \geq 10$ y $n(1 - p) \geq 10$], tanto X como el estimador $\hat{p} = X/n$ están normalmente distribuidos en forma aproximada. Primero se consideran pruebas con muestras grandes basadas en este último hecho y luego se acude al caso de muestra pequeña que usa de modo directo la distribución binomial.

Pruebas con muestra grande

Las pruebas con muestra grande relacionadas con p son un caso especial de los procedimientos con muestra grande para un parámetro θ. Sea $\hat{\theta}$ un estimador de θ que es (por lo menos de manera aproximada) insesgado y que tiene aproximadamente una distribución normal. La hipótesis nula tiene la forma H_0: $\theta = \theta_0$, donde θ_0 denota un número (el valor nulo) apropiado al contexto del problema. Suponga que cuando H_0 es verdadera, la desviación estándar de $\hat{\theta}$, $\sigma_{\hat{\theta}}$, no implica parámetros desconocidos. Por ejemplo, si $\theta = \mu$ y $\hat{\theta} = \overline{X}$, $\sigma_{\hat{\theta}} = \sigma_{\overline{X}} = \sigma/\sqrt{n}$, la cual no implica parámetros desconocidos sólo si se conoce el valor de σ. Al estandarizar $\hat{\theta}$ conforme a la suposición de que H_0 es verdadera (de modo que $E(\hat{\theta}) = \theta_0$) se obtiene un estadístico de prueba para muestra grande:

$$\text{Estadístico de prueba: } Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$$

Si la hipótesis alternativa es H_a: $\theta > \theta_0$, la región de rechazo $z \geq z_\alpha$ especifica una prueba de cola superior cuyo nivel significativo es aproximadamente α. Las otras dos alternativas, H_a: $\theta < \theta_0$ y H_a: $\theta \neq \theta_0$, se someten a una prueba z de cola inferior y a una prueba z de dos colas, respectivamente.

En el caso $\theta = p$, $\sigma_{\hat{\theta}}$ no implicará parámetros desconocidos cuando H_0 es verdadera, aunque esto es atípico. Cuando $\sigma_{\hat{\theta}}$ implica parámetros desconocidos, a menudo es posible utilizar una desviación estándar estimada $S_{\hat{\theta}}$ en lugar de $\sigma_{\hat{\theta}}$ y seguir teniendo Z aproximadamente distribuida de manera normal cuando H_0 es verdadera (porque cuando n es grande $s_{\hat{\theta}} \approx \sigma_{\hat{\theta}}$ para la mayoría de las muestras). La prueba con muestra grande de la sección previa da un ejemplo de esto, como σ casi siempre es desconocida, se utiliza $s_{\hat{\theta}} = s_{\bar{X}} = s/\sqrt{n}$ en lugar de $\sigma/\sqrt{n}$ en el denominador de z.

El estimador $\hat{p} = X/n$ es ($E(\hat{p}) = p$) insesgado y su distribución es aproximadamente normal y su desviación estándar es $\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$. Estos hechos se utilizaron en la sección 7.2 para obtener un intervalo de confianza para p. Cuando H_0 es verdadera, $E(\hat{p}) = p_0$ y $\sigma_{\hat{p}} = \sqrt{p_0(1-p_0)/n}$, $\sigma_{\hat{p}}$ no implica parámetros desconocidos. Se concluye entonces que cuando n es grande y H_0 es verdadera, el estadístico de prueba

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

tiene aproximadamente una distribución normal estándar. Si la hipótesis alternativa es H_a: $p > p_0$ y se utiliza la región de rechazo de cola superior $z \geq z_\alpha$, entonces

$$\begin{aligned} P(\text{error de tipo I}) &= P(H_0 \text{ es rechazada cuando es verdadera}) \\ &= P(Z \geq z_\alpha \text{ cuando } Z \text{ tiene aproximadamente una distribución normal estándar}) \approx \alpha \end{aligned}$$

Por consiguiente, el nivel de significancia deseado α se obtiene utilizando el valor crítico que capture el área α en la cola superior de la curva z. Las regiones de rechazo para las otras dos hipótesis alternativas, cola inferior para H_a: $p < p_0$ y dos colas para H_a: $p \neq p_0$ se justifican de manera análoga.

Hipótesis nula: H_0: $p = p_0$

Valor del estadístico de prueba: $z = \dfrac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$

Hipótesis alternativa	Región de rechazo
H_a: $p > p_0$	$z \geq z_\alpha$ (cola superior)
H_a: $p < p_0$	$z \leq -z_\alpha$ (cola inferior)
H_a: $p \neq p_0$	$z \geq z_{\alpha/2}$ o $z \leq -z_{\alpha/2}$ (dos colas)

Estos procedimientos de prueba son válidos siempre que $np_0 \geq 10$ y $n(1-p_0) \geq 10$.

Ejemplo 8.11 El corcho natural en botellas de vino está sujeto a deterioro y como resultado el vino de esas botellas puede experimentar contaminación. El artículo "Effects of Bottle Closure Type on Consumer Perceptions of Wine Quality" (*Amer. J. of Enology and Viticulture*, 2007: 182–191) informó que, en una degustación de chardonnay comerciales, 16 de 91 botellas se consideraron echadas a perder en cierta medida por las características asociadas del corcho. ¿Estos datos proporcionan una fuerte evidencia para concluir que más del 15% de todas las botellas están contaminadas de esta manera? Se realizará una prueba de hipótesis con un nivel de significancia de .10.

1. p = la verdadera proporción de todas las botellas de chardonnay comerciales consideradas inservibles en cierta medida por las características asociadas del corcho.
2. La hipótesis nula es H_0: $p = .15$.
3. La hipótesis alternativa es H_a: $p > .15$, la afirmación de que el porcentaje de población supera el 15%.

4. Como $np_0 = 91(.15) = 13.65 > 10$ y $nq_0 = 91(.85) = 77.35 > 10$, la prueba z con muestra grande puede ser utilizada. El valor estadístico de prueba es $z = (\hat{p} - .15)/\sqrt{(.15)(.85)/n}$.
5. La forma de H_a implica que una prueba de cola superior es apropiada: rechazar H_0 si $z \geq z_{.10} = 1.28$.
6. $\hat{p} = 16/91 = .1758$, de donde se obtiene
$$z = (.1758 - .15)/\sqrt{(.15)(.85)/91} = .0258/.0374 = .69$$
7. Como $.69 < 1.28$, z no está en la región de rechazo. Al nivel de significancia .10, la hipótesis nula no puede rechazarse. Aunque el porcentaje de botellas contaminadas en la muestra es un poco superior al 15%, el porcentaje de la muestra no es lo suficientemente grande como para concluir que el porcentaje de la población supera el 15%. La diferencia entre la proporción de la muestra .1758 y el valor nulo .15 puede ser explicado adecuadamente por la variabilidad de muestreo. ■

β y determinación del tamaño de la muestra Cuando H_0 es verdadera, el estadístico de prueba Z tiene aproximadamente una distribución normal estándar. Supóngase ahora que H_0 *no* es verdadera y que $p = p'$. Entonces Z sigue teniendo aproximadamente una distribución normal (porque es una función lineal de $\hat{p}$), pero su valor medio y su varianza ya no son 0 y 1, respectivamente. En su lugar,

$$E(Z) = \frac{p' - p_0}{\sqrt{p_0(1 - p_0)/n}} \qquad V(Z) = \frac{p'(1 - p')/n}{p_0(1 - p_0)/n}$$

La probabilidad de un error de tipo II para una prueba de cola superior es $\beta(p') = P(Z > z_\alpha$ cuando $p = p')$. Esto se puede calcular usando la media y la varianza dadas para estandarizar y entonces referirse a la función de distribución acumulativa normal estándar. Además, si se desea que el nivel de prueba α también tenga $\beta(p') = \beta$ para un valor especificado de β, esta ecuación se resuelve para la n necesaria como en la sección 8.2. En el recuadro adjunto se dan expresiones generales para $\beta(p')$ y n.

Hipótesis alternativa	$\beta(p')$
H_a: $p > p_0$	$\Phi\left[\dfrac{p_0 - p' + z_\alpha\sqrt{p_0(1 - p_0)/n}}{\sqrt{p'(1 - p')/n}}\right]$
H_a: $p < p_0$	$1 - \Phi\left[\dfrac{p_0 - p' - z_\alpha\sqrt{p_0(1 - p_0)/n}}{\sqrt{p'(1 - p')/n}}\right]$
H_a: $p \neq p_0$	$\Phi\left[\dfrac{p_0 - p' + z_{\alpha/2}\sqrt{p_0(1 - p_0)/n}}{\sqrt{p'(1 - p')/n}}\right]$ $-\Phi\left[\dfrac{p_0 - p' - z_{\alpha/2}\sqrt{p_0(1 - p_0)/n}}{\sqrt{p'(1 - p')/n}}\right]$

El tamaño de muestra n con el cual la prueba de nivel α también satisface $\beta(p') = \beta$ es

$$n = \begin{cases} \left[\dfrac{z_\alpha\sqrt{p_0(1 - p_0)} + z_\beta\sqrt{p'(1 - p')}}{p' - p_0}\right]^2 & \text{prueba de una cola} \\ \left[\dfrac{z_{\alpha/2}\sqrt{p_0(1 - p_0)} + z_\beta\sqrt{p'(1 - p')}}{p' - p_0}\right]^2 & \text{prueba de dos colas (una solución aproximada)} \end{cases}$$

Ejemplo 8.12 Un servicio de mensajería anuncia que por lo menos 90% de todos los paquetes llevados a su oficina alrededor de las 9 a.m. para entrega en la misma ciudad son entregados alrededor del mediodía de ese mismo día. Sea p la proporción verdadera de dichos paquetes que son entregados como se anuncia y considere las hipótesis H_0: $p = .9$ contra H_a: $p < .9$. Si sólo 80% de los paquetes son entregados como se anuncia, ¿qué tan probable es que una prueba de nivel .01 basada en $n = 225$ paquetes detectará tal alejamiento de H_0? ¿Cuál debe ser el tamaño de la muestra para garantizar que $\beta(.8) = .01$? Con $\alpha = .01$, $p_0 = .9$, $p' = .8$ y $n = 225$,

$$\begin{aligned}\beta(.8) &= 1 - \Phi\left(\frac{.9 - .8 - 2.33\sqrt{(.9)(.1)/225}}{\sqrt{(.8)(.2)/225}}\right)\\ &= 1 - \Phi(2.00) = .0228\end{aligned}$$

Así pues la probabilidad de que H_0 sea rechazada si se realiza la prueba cuando $p = .8$ es .9772, aproximadamente 98% de todas las muestras darán por resultado el rechazo correcto de H_0.

Con $z_\alpha = z_\beta = 2.33$ en la fórmula del tamaño de la muestra se obtiene

$$n = \left[\frac{2.33\sqrt{(.9)(.1)} + 2.33\sqrt{(.8)(.2)}}{.8 - .9}\right]^2 \approx 266$$ ■

Pruebas con muestra pequeña

Los procedimientos de prueba cuando el tamaño de muestra n es pequeño están basados directamente en la distribución binomial en lugar de en la aproximación normal. Considérese la hipótesis alternativa H_a: $p > p_0$ y de nuevo sea X el número de éxitos en la muestra. Entonces X es el estadístico de prueba y la región de rechazo de cola superior tiene la forma $x \geq c$. Cuando H_0 es verdadera, X tiene una distribución binomial con parámetros n y p_0, por lo tanto

$$\begin{aligned}P(\text{error de tipo I}) &= P(H_0 \text{ es rechazada cuando es verdadera})\\ &= P(X \geq c \text{ cuando } X \sim \text{Bin}(n, p_0))\\ &= 1 - P(X \leq c - 1 \text{ cuando } X \sim \text{Bin}(n, p_0)\\ &= 1 - B(c - 1; n, p_0)\end{aligned}$$

A medida que el valor crítico c disminuye, más valores x están incluidos en la región de rechazo y P(error de tipo I) se incrementa. Como X tiene una distribución de probabilidad discreta, por lo general no es posible hallar un valor de c con el cual P(error de tipo I) sea exactamente el nivel de significancia α deseado (p. ej., .05 o .01). En su lugar, se utiliza la región de rechazo más grande de la forma $\{c, c + 1, \ldots, n\}$ que satisface $1 - B(c - 1; n, p_0) \leq \alpha$.

Sea p' un valor alternativo de $p(p' > p_0)$. Cuando $p = p'$, $X \sim \text{Bin}(n, p')$, por lo tanto

$$\begin{aligned}\beta(p') &= P(\text{error de tipo II cuando } p = p')\\ &= P(X < c \text{ cuando } X \sim \text{Bin}(n, p')) = B(c - 1; n, p')\end{aligned}$$

Es decir, $\beta(p')$ es el resultado de un cálculo de probabilidad binomial directo. El tamaño de muestra n necesario para garantizar que una prueba de nivel α tiene una β especificada con valor alternativo particular p' debe ser determinado mediante ensayo y error utilizando la función de distribución acumulativa binomial.

Los procedimientos de prueba para H_a: $p < p_0$ y para H_a: $p \neq p_0$ se construyen de manera similar. En el primer caso, la región de rechazo apropiada tiene la forma $x \leq c$ (una prueba de cola inferior). El valor crítico c es el número más grande que satisface $B(c; n, p_0) \leq \alpha$. La región de rechazo cuando la hipótesis alternativa es H_a: $p \neq p_0$ se compone tanto de valores x grandes como pequeños.

Ejemplo 8.13 Un fabricante de plástico desarrolló un nuevo tipo de bote de plástico para la basura y propone venderlo con una garantía incondicional de 6 años. Para ver si esto es factible desde el punto de vista económico, 20 botes prototipo se someten a una prueba acelerada de duración para simular 6 años de uso. La garantía propuesta se modificará sólo si los datos muestrales sugieren fuertemente que menos del 90% de los botes sobrevivirían el periodo de 6 años. Sea p la proporción de todos los botes que sobreviven la prueba acelerada. Las hipótesis pertinentes son H_0: $p = .9$ contra H_a: $p < .9$. La decisión se basará en el estadístico de prueba X, el número de entre los 20 que sobreviven. Si el nivel de significancia deseado es $\alpha = .05$, c debe satisfacer $B(c; 20, .9) \leq .05$. De acuerdo con la tabla A.1 del apéndice, $B(15; 20, .9) = .043$, mientras que $B(16; 20, .9) = .133$. La región de rechazo apropiada es por consiguiente $x \leq 15$. Si la prueba acelerada da por resultado $x = 14$, H_0 sería rechazada a favor de H_a y se modificaría la garantía propuesta. La probabilidad de un error de tipo II con el valor alternativo $p' = .8$ es

$$\begin{aligned}\beta(.8) &= P(H_0 \text{ no es rechazada cuando } X \sim \text{Bin}(20, .8)) \\ &= P(X \geq 16 \text{ cuando } X \sim \text{Bin}(20, .8)) \\ &= 1 - B(15; 20, .8) = 1 - .370 = .630\end{aligned}$$

Es decir, cuando $p = .8$, 63% de todas las muestras compuestas de $n = 20$ botes darían por resultado que H_0 sea incorrectamente no rechazada. Esta probabilidad de error es alta porque 20 es un tamaño de muestra pequeño y $p' = .8$ se acerca al valor nulo $p_0 = .9$. ■

EJERCICIOS Sección 8.3 (37–46)

37. Una caracterización común de las personas obesas es que su índice de masa corporal es de al menos 30 [IMC = peso/(altura)^2, donde la altura está en metros y el peso en kilogramos]. El artículo "The Impact of Obesity on Illness Absence and Productivity in an Industrial Population of Petrochemical Workers" (*Annals of Epidemiology*, 2008: 8−14) informó que en una muestra de mujeres trabajadoras, 262 tenían un índice de masa corporal inferior a 25, 159 tenían IMC que era al menos 25 pero no más de 30 y 120 tenían un IMC superior a 30. ¿Hay pruebas contundentes para concluir que más del 20% de individuos en la población analizada son obesos?

- **a.** Establezca las hipótesis de prueba adecuadas utilizando el enfoque de región de rechazo con un nivel de significancia de .05.
- **b.** Explique en el contexto de este escenario qué constituye errores de tipos I y II.
- **c.** ¿Cuál es la probabilidad de no concluir que más del 20% de la población es obesa cuando el porcentaje real de los individuos obesos es del 25%?

38. Un fabricante de baterías de níquel-hidrógeno selecciona al azar 100 placas de níquel para probar las celdas, someterlas a ciclos un número especificado de veces y concluye que 14 de ellas se ampollan en tales circunstancias.

- **a.** ¿Proporciona esto una evidencia precisa para concluir que más de 10% de todas las placas se ampollan en tales circunstancias? Formule y pruebe las hipótesis apropiadas con un nivel de significancia de .05. Al llegar a su conclusión, ¿qué tipo de error pudo haber cometido?
- **b.** Si es realmente el caso que 15% de todas las placas se ampollan en estas circunstancias y se utiliza un tamaño de muestra de 100, ¿qué tan probable es que la hipótesis nula del inciso (a) no sea rechazada por la prueba de nivel .05? Responda esta pregunta para un tamaño de muestra de 200.
- **c.** ¿Cuántas placas tendrían que ser probadas para tener $\beta(.15) = .10$ para la prueba del inciso (a)?

39. Una muestra aleatoria de 150 donaciones recientes en un banco de sangre revela que 82 fueron de sangre tipo A. ¿Sugiere esto que el porcentaje real de donaciones tipo A difiere de 40%, el porcentaje de la población que tiene sangre tipo A? Realice una prueba de las hipótesis apropiadas utilizando un nivel de significancia de .01. ¿Habría sido diferente su conclusión si se hubiera utilizado un nivel de significancia de .05?

40. Se sabe que aproximadamente 2/3 de todos los seres humanos tienen un ojo o pie derecho dominantes. ¿Existe también dominio del lado derecho en el acto de besar? El artículo "Human Behavior: Adult Persistence of Head-Turning Asymmetry" (*Nature*, 2003: 771) reportó que en una muestra aleatoria de 124 parejas que se besan, ambas personas en 80 de las parejas tendieron a inclinarse más hacia la derecha que hacia la izquierda.

- **a.** Si 2/3 de las parejas que se besan exhiben esta tendencia de inclinarse hacia la derecha, ¿cuál es la probabilidad de que el número en una muestra de 124 que lo hacen así difiera del valor esperado en por lo menos lo que en realidad se observó?
- **b.** ¿Sugiere este resultado del experimento que la cifra de 2/3 es no factible como comportamiento al besarse? Formule y pruebe las hipótesis apropiadas.

41. El artículo referido en el ejemplo 8.11 también informó que en una muestra de 106 consumidores de vino, 22 (20.8%) opina que las tapas de tornillo son un sustituto aceptable para los corchos naturales. Suponga que una bodega particular decide utilizar tapas de rosca en uno de sus vinos a menos que haya

pruebas sólidas que sugieran que un porcentaje inferior al 25% de los consumidores de vino encuentran esto aceptable.

a. Utilizando un nivel de significancia de .10, ¿qué le recomendaría a la bodega?

b. Para la hipótesis probada en el inciso (a), describa en el contexto qué tipos de errores I y II serían y diga qué tipo de error pudo haberse cometido.

42. Con las fuentes locales de materiales de construcción disminuyendo desde hace varios años, alrededor de 60,000 casas fueron construidas con paneles de yeso chinos importados. De acuerdo con el artículo "Report Links Chinese Drywall to Home Problems" (*New York Times*, 24 de nov. de 2009), los investigadores federales identificaron una fuerte asociación entre los productos químicos en los paneles de yeso y problemas eléctricos, y también una fuerte evidencia de dificultades respiratorias debidas a la emisión de gases de sulfuro de hidrógeno. Un examen exhaustivo de 51 casas encontró que 41 tenían este tipo de problemas. Suponga que estas 51 fueron muestreadas al azar de la población de todas las casas que tienen paneles de yeso chinos.

a. ¿Los datos proporcionan una fuerte evidencia para concluir que más del 50% de todas las casas con paneles de yeso chinos tienen problemas eléctricos o ambientales? Realice una prueba de hipótesis usando $\alpha = .01$.

b. Calcule un límite de confianza inferior con un nivel de confianza del 99% para el porcentaje de todos los hogares que tengan problemas eléctricos o ambientales.

c. Si en realidad es el caso que el 80% de todas las casas tienen tales problemas, ¿qué tan probable es que la prueba del inciso (a) no llegue a la conclusión de que más del 50% los tiene?

43. Una aerolínea desarrolló un club de viajeros ejecutivos sobre la premisa de que 5% de sus clientes actuales satisfacen los requisitos para la membresía. Una muestra aleatoria de 500 clientes arrojó 40 que calificarían.

a. Con estos datos, pruebe a un nivel de .01 la hipótesis nula de que la premisa de la compañía es correcta contra la alternativa de que no es correcta.

b. ¿Cuál es la probabilidad de que cuando se utiliza la prueba del inciso (a), la premisa de la compañía será juzgada correcta cuando en realidad 10% de todos los clientes actuales cumplan los requisitos?

44. Cada uno de un grupo de 20 tenistas intermedios recibe dos raquetas, una con cuerdas de nailon y la otra con cuerdas de tripa sintética. Tras varias semanas de jugar con las dos raquetas, a cada jugador se le pide que manifieste una preferencia por uno de los dos tipos de cuerdas. Sea p la proporción de todos los jugadores que preferirían las de tripa y sea X el número de jugadores en la muestra que prefieren las de tripa. Como las cuerdas de tripa son más caras, considere la hipótesis nula de que cuando mucho 50% de todos los jugadores prefieren las cuerdas de tripa. Se simplifica esto a H_0: $p = .5$, planeando rechazar H_0 sólo si la evidencia muestral favorece fuertemente las cuerdas de tripa.

a. ¿Cuál de las regiones de rechazo {15, 16, 17, 18, 19, 20}, {0, 1, 2, 3, 4, 5} o {0, 1, 2, 3, 17, 18, 19, 20} es más apropiada y por qué las otras dos no son apropiadas?

b. ¿Cuál es la probabilidad de un error de tipo I para la región seleccionada del inciso (a)? ¿Especifica la región una prueba de nivel .05? ¿La prueba de nivel .05 es la mejor?

c. Si 60% de todos los fanáticos prefieren las cuerdas de tripa, calcule la probabilidad de un error de tipo II utilizando la región apropiada del inciso (a). Repita si 80% de todos los fanáticos prefieren las cuerdas de tripa.

d. Si 13 de los 20 jugadores prefieren las cuerdas de tripa, ¿deberá ser rechazada H_0 si se utiliza un nivel de significancia de .10?

45. Un fabricante de artículos de plomería creó un nuevo tipo de llave de agua sin empaques. Sea $p = P$(en una llave seleccionada al azar de este tipo aparecerá una fuga dentro de 2 años de uso normal). El fabricante decidió proseguir con la producción a menos que se pueda determinar que p es demasiado grande; el valor límite aceptable de p se especifica como .10. El fabricante decide someter a n de estas llaves a una prueba acelerada (simulando de manera aproximada dos años de uso normal). Con $X =$ el número entre las n llaves en las que aparecen fugas antes de que concluya la prueba, la producción arrancará a menos que la X observada sea demasiado grande. Se decidió que si $p = .10$, la probabilidad de no proseguir deberá ser cuando mucho de .10, en tanto que si $p = .30$ la probabilidad de proseguir deberá ser cuando mucho de .10. ¿Se puede utilizar $n = 10$?, ¿$n = 20$?, ¿$n = 25$? ¿Cuál es la región de rechazo apropiada con la n seleccionada y cuáles son las probabilidades de error reales cuando se utiliza esta región?

46. Científicos piensan que los robots desempeñarán un papel crucial en las fábricas en las siguientes décadas. Suponga que en un experimento para determinar si el uso de robots para instalar cables de computadora es factible, se utilizó un robot para ensamblar 500 cables. Se examinaron los cables y se encontraron 15 defectuosos. Si los ensambladores humanos tienen una proporción de cables defectuosos de .035 (3.5%), ¿apoyan estos datos la hipótesis de que la proporción de cables defectuosos es menor con robots que con humanos? Use un nivel de significancia de .01.

8.4 Valores *P*

Utilizar el método de región de rechazo para poner a prueba hipótesis, implica seleccionar primero un nivel α de significancia. Luego, después de calcular el valor del estadístico de prueba, la hipótesis nula H_0 se rechaza si el valor cae en la región de rechazo, en caso contrario no. Considérese ahora otra forma de llegar a una conclusión en un análisis de las pruebas de hipótesis. Este enfoque alternativo se basa en el cálculo de una probabilidad específica denominada *valor P*. Una ventaja es que el valor P proporciona una medida intuitiva de la fuerza de la evidencia en los datos en contra de H_0.

DEFINICIÓN

El **valor *P*** es la probabilidad, calculada suponiendo que la hipótesis nula es cierta, de obtener un valor del estadístico de prueba por lo menos tan contradictorio para H_0 como el valor calculado a partir de la muestra disponible.

Esta definición es importante. Los siguientes son algunos puntos clave:

- El valor *P* es una probabilidad.
- Esta probabilidad se calcula suponiendo que la hipótesis nula es cierta.
- ¡Tenga cuidado: el valor *P* no es la probabilidad de que H_0 sea cierta, ni es una probabilidad de error!
- Para determinar el valor *P*, primero se debe decidir qué valores del estadístico de prueba son al menos tan contradictorios para H_0 como el valor obtenido de la muestra.

Ejemplo 8.14 Aguas pluviales urbanas pueden estar contaminadas por muchas fuentes, incluyendo las pilas desechadas. Cuando se rompen, estas baterías liberan metales de importancia medioambiental. El artículo "Urban Battery Litter" (*J. of Environ. Engr.*, 2009: 46−57), presentó los datos que resumen las características de una variedad de baterías que se encuentran en las zonas urbanas de Cleveland. Una muestra de 51 baterías Panasonic AAA dio una media muestral de masa de zinc de 2.06 gramos y una desviación estándar de la muestra de .141 g. ¿Estos datos proporcionan pruebas convincentes para concluir que la media de la población de la masa de zinc sea superior a 2.0 g?

Con μ se denota el promedio real de la masa de zinc de este tipo de baterías, las hipótesis relevantes son H_0: $\mu = 2.0$ contra H_a: $\mu > 2.0$. El tamaño de la muestra es suficientemente grande para que la prueba *z* se pueda utilizar sin hacer ninguna suposición específica sobre la forma de la distribución de la población. El valor del estadístico de prueba es

$$z = \frac{\bar{x} - 2.0}{s/\sqrt{n}} = \frac{2.06 - 2.0}{.141/\sqrt{51}} = 3.04$$

Ahora debe decidirse qué valores de *z* son al menos tan contradictorios para H_0. Se considera primero una tarea fácil: ¿qué valores de $\bar{x}$ son por lo menos tan contradictorios para la hipótesis nula como 2.06, la media de las observaciones de la muestra? Debido a que el símbolo $>$ aparece en H_a, debe quedar claro que 2.10 es al menos tan contradictorio para H_0 como 2.06, por lo que en realidad es *cualquier* valor $\bar{x}$ que supere los 2.06. Sin embargo, un valor $\bar{x}$ que supera los 2.06 corresponde a un valor de *z* que excede 3.04. Así, el valor *P* es

$$\text{Valor } P = P(Z \geq 3.04 \text{ cuando } \mu = 2.0)$$

Dado que el estadístico de prueba *Z* fue creado al restar el valor nulo 2.0 en el numerador, cuando $\mu = 2.0$; es decir, cuando H_0 es cierta, *Z* tiene aproximadamente una distribución normal estándar. Como consecuencia de ello,

$$\begin{aligned}\text{Valor } P &= P(Z \geq 3.04 \text{ cuando } \mu = 2.0) \approx \text{área bajo la curva } z \text{ a la derecha de } 3.04 \\ &= 1 - \Phi(3.04) = .0012\end{aligned}$$

■

En breve se ilustra cómo determinar el valor *P* para cualquier prueba *z* o *t*; es decir, cualquier prueba en la que la distribución de referencia es la distribución normal estándar (y la curva *z*) o alguna distribución *t* (y la curva *t* correspondiente). Por el momento, sin embargo, lo más importante será llegar a una conclusión una vez que el valor de *P* está disponible. Debido a que es una probabilidad, el valor *P* debe estar entre 0 y 1. ¿Qué tipos de valores *P* aportan pruebas en contra de la hipótesis nula? Se consideran dos casos concretos:

- Valor $P = .250$: en este caso, el 25% de todos los valores posibles del estadístico de prueba son al menos tan contradictorios para H_0 como el que salió de la muestra. Así que los datos no son tan contradictorios para la hipótesis nula.
- Valor $P = .0018$: aquí, sólo .18% (mucho menos del 1%) de todos los valores posibles del estadístico de prueba son al menos tan contradictorios para H_0 como el que se obtuvo. Así, la muestra parece ser muy contradictoria para la hipótesis nula.

De manera más general, *cuanto menor sea el valor de P, es mayor la evidencia que hay en los datos de la muestra en contra de la hipótesis nula y la hipótesis alternativa*. Es decir, H_0 debe ser rechazada a favor de H_a, cuando el valor P es suficientemente pequeño. Pero, ¿qué es "suficientemente pequeño"?

Regla de decisión basada en el valor *P*

Se selecciona un nivel de significancia α (como antes, el tipo I de error deseado para la probabilidad). A continuación,

rechazar H_0 si el valor $P \leq \alpha$

no rechazar H_0 si el valor $P > \alpha$

De esta manera, si el valor P excede el nivel de significancia elegido, la hipótesis nula no se puede rechazar a ese nivel. Pero si el valor P es igual o menor que α, entonces hay pruebas suficientes para justificar el rechazo de H_0. En el ejemplo 8.14 se calculó el valor $P = .0012$. Luego, utilizando un nivel de significancia de .01, se rechazaría la hipótesis nula a favor de la hipótesis alternativa porque $.0012 \leq .01$. Sin embargo, supóngase que se selecciona un nivel de significancia de tan sólo .001, lo que requiere pruebas más significativas de los datos antes de que se rechace H_0. En este caso no se rechaza H_0 porque $.0012 > .001$

¿Cómo funciona la regla de decisión basada en el valor P en comparación con la regla de decisión empleada en el método de región de rechazo? *Los dos procedimientos, el método de región de rechazo y el método del valor P, en realidad son idénticos*. Cualquiera que sea la conclusión alcanzada empleando el enfoque de región de rechazo con un α particular, se llega a la misma conclusión por medio del valor P con el mismo α.

Ejemplo 8.15 El problema del contenido de nicotina analizado en el ejemplo 8.5 involucra probar H_0: $\mu = 1.5$ contra H_a: $\mu > 1.5$ utilizando la prueba z (es decir, una prueba que utiliza la curva z como la distribución de referencia). La desigualdad en H_a implica que la región de rechazo de cola superior $z \geq z_\alpha$ es la adecuada. Supóngase que $z = 2.10$. Luego, utilizando exactamente el mismo razonamiento que en el ejemplo 8.14 se tiene que valor $P = 1 - \Phi(2.10) = .0179$. Se consideran ahora las pruebas con varios niveles de significancia diferentes:

$$\alpha = .10 \Rightarrow z_\alpha = z_{.10} = 1.28 \Rightarrow 2.10 \geq 1.28 \Rightarrow \text{rechazar } H_0$$
$$\alpha = .05 \Rightarrow z_\alpha = z_{.05} = 1.645 \Rightarrow 2.10 \geq 1.645 \Rightarrow \text{rechazar } H_0$$
$$\alpha = .01 \Rightarrow z_\alpha = z_{.01} = 2.33 \Rightarrow 2.10 < 2.33 \Rightarrow \text{no rechazar } H_0$$

Como el valor $P = .0179 \leq .10$ y $.0179 \leq .05$, se utiliza el enfoque de resultados de valor P en el rechazo de H_0 para los dos primeros niveles de significancia. Sin embargo, para $\alpha = .01$, 2.10 no se encuentra en la región de rechazo y .0179 es mayor que .01. De manera más general, siempre que α es menor que el valor P de .0179, el valor crítico z_α estará más allá del valor calculado de z y H_0 no puede ser rechazada por cualquiera de los métodos. Esto se ilustra en la figura 8.7.

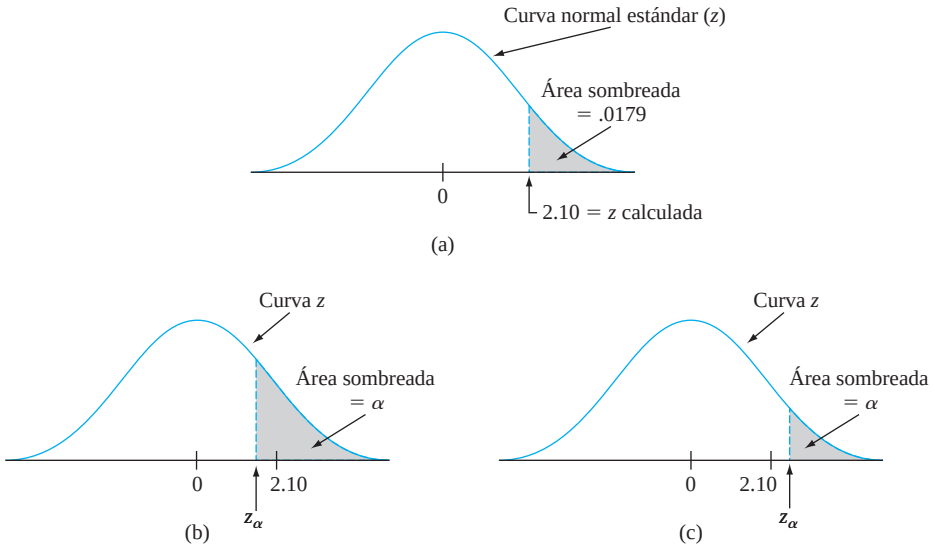


Figura 8.7 Relación entre α y área de cola capturada al calcular z: (a) área de cola capturada al calcular z; (b) cuando $\alpha < .0179$, $z_\alpha < 2.10$ y H_0 es rechazada; (c) cuando $\alpha < .0179$, $z_\alpha > 2.10$ y H_0 no es rechazada

Reconsidérese una vez más el valor $P = .0012$ en el ejemplo 8.14. H_0 puede ser rechazada sólo si $.0012 \leq \alpha$. Así, la hipótesis nula puede ser rechazada si $\alpha = .05$ o $.01$ o $.005$ o $.0015$ o $.00125$. ¿Cuál es el *menor* nivel α de significancia aquí para el que H_0 puede ser rechazada? Es el valor $P = .0012$.

PROPOSICIÓN:

El valor P es el nivel de significancia α más pequeño para el cual la hipótesis nula puede ser rechazada. Debido a esto, el valor P es tomado como alternativa para el **nivel de significancia observado** para los datos.

Se acostumbra llamar *significativos* a los datos cuando H_0 es rechazada y *no significativos* de lo contrario. El valor P es entonces el nivel más pequeño al cual los datos son significativos. Una manera fácil de visualizar la comparación del valor P con el nivel α seleccionado es trazar una imagen como la de la figura 8.8. El cálculo del valor P depende de si la prueba es de cola superior, inferior o de dos colas. No obstante, una vez calculada, la comparación con α no depende de qué tipo de prueba se utilizó.

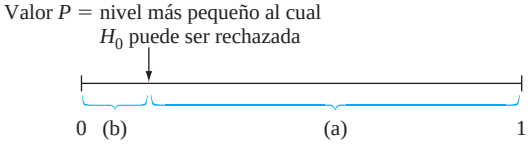


Figura 8.8 Comparación de α y el valor P: (a) rechazar H_0 cuando α queda aquí; (b) no rechazar H_0 cuando α queda aquí

Ejemplo 8.16 El verdadero tiempo promedio para el alivio inicial del dolor con el analgésico más vendido es de 10 minutos. Sea μ que denota el tiempo promedio real de alivio para un nuevo medicamento desarrollado por la empresa. Supóngase que cuando se analizan los datos de un experimento con el analgésico nuevo, el valor P para probar H_0: $\mu = 10$ frente a H_a:

$\mu < 10$ se calcula como .0384. Dado que $\alpha = .05$ es mayor que el valor P [.05 se encuentra en el intervalo (a) de la figura 8.8], H_0 sería rechazada por toda persona que efectúe la prueba a nivel de .05. Sin embargo, a nivel de .01, H_0 no se rechaza porque .01 es menor que el nivel más bajo (.0384) en el que se puede rechazar H_0. ■

Los paquetes de cómputo para estadística más utilizados incluyen automáticamente un valor de P cuando se lleva a cabo un análisis de prueba de hipótesis. Puede extraerse una conclusión directamente de la salida, sin hacer referencia a una tabla de valores críticos. Con el valor P en la mano, un investigador puede ver rápidamente para qué niveles de significancia H_0 podría ser o no rechazada. Además, cada individuo puede seleccionar su nivel de significancia propio. Adicionalmente, conocer el valor P permite que alguien que toma decisiones distinga entre una llamada cercana (por ejemplo, $\alpha = .05$, valor $P = .0498$) y una conclusión muy bien definida (por ejemplo, $\alpha = .05$, valor $P = .0003$), algo que no sería posible sólo a partir de la declaración "se puede rechazar H_0 al nivel de significancia .05".

Valores *P* para pruebas *z*

El valor P para una prueba z (una basada en un estadístico de prueba cuya distribución cuando H_0 es verdadera es por lo menos aproximadamente normal estándar) es fácil de determinar a partir de la información de la tabla A.3 del apéndice. Considérese una prueba de cola superior y sea z el valor calculado del estadístico de prueba Z. La hipótesis nula es rechazada si $z \geq z_\alpha$ y el valor P es el α más pequeño con el cual éste es el caso. Como z_α se incrementa a medida que α disminuye, el valor P es el valor de α con el cual $z = z_\alpha$. Es decir, el valor P es exactamente el área capturada por el valor z calculado en la cola superior de la curva normal estándar. El área acumulativa correspondiente es $\Phi(z)$, así que en este caso el valor $P = 1 - \Phi(z)$.

Un argumento análogo para una prueba de cola inferior demuestra que el valor P es el área capturada por el valor z calculado en la cola inferior de la curva normal estándar. Se debe tener más cuidado en el caso de una prueba de dos colas. Supóngase primero que z es positivo. Entonces el valor P es el valor de α que satisface $z = z_{\alpha/2}$ (es decir, z calculado = valor crítico de cola superior). Esto dice que el área capturada en la cola superior es la mitad del valor P, de modo que valor $P = 2[1 - \Phi(z)]$. Si z es negativo, el valor P es el α con el cual $z = -z_{\alpha/2}$ o, de forma equivalente, $-z = z_{\alpha/2}$, así que valor $P = 2[1 - \Phi(-z)]$. Como $-z = |z|$ cuando z es negativo, valor $P = 2[1 - \Phi(|z|)]$ con z positivo o negativo.

$$\text{valor } P\text{: } P = \begin{cases} 1 - \Phi(z) & \text{para una prueba } z \text{ de cola superior} \\ \Phi(z) & \text{o una prueba } z \text{ de cola inferior} \\ 2[1 - \Phi(|z|)] & \text{para una prueba } z \text{ de dos colas} \end{cases}$$

Cada una de éstas es la probabilidad de tener un valor por lo menos tan extremo como el que se obtuvo (suponiendo que H_0 es verdadera). Los tres casos se ilustran en la figura 8.9.

El siguiente ejemplo ilustra el uso del método del valor P para la prueba de hipótesis por medio de una secuencia de pasos modificados con respecto a la secuencia previamente recomendada.

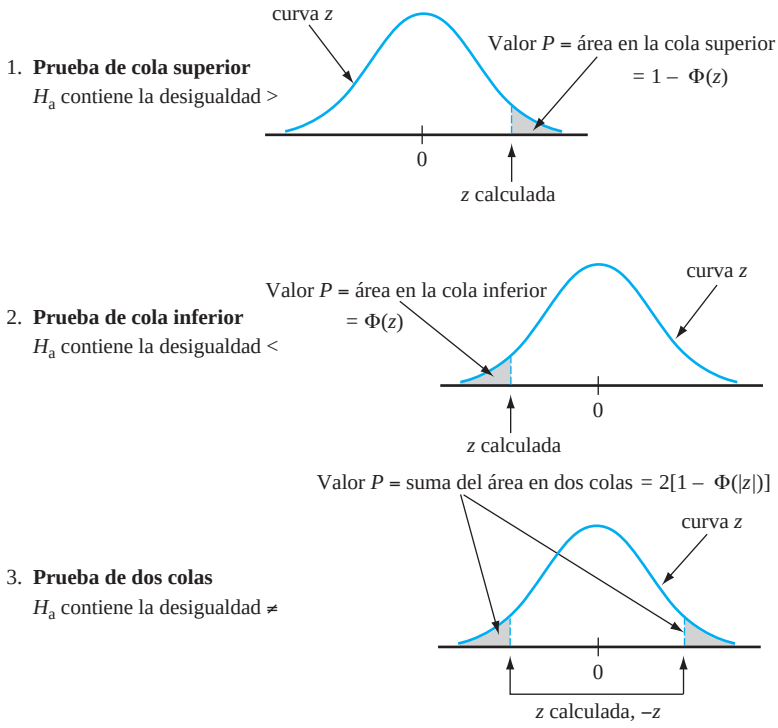


Figura 8.9 Determinación del valor *P* para una prueba *z*

Ejemplo 8.17 El espesor deseado de obleas de silicio utilizadas en cierto tipo de circuito integrado es de 245 μm. Se obtiene una muestra de 50 obleas y se determina el espesor de cada una, dando como resultado un espesor medio de la muestra de 246.10 μm y una desviación estándar muestral de 3.60 μm. ¿Sugieren estos datos que el espesor de oblea promedio verdadero es algún otro diferente del valor deseado?

1. Parámetro de interés: μ = espesor de oblea promedio verdadero

2. Hipótesis nula: H_0: $\mu = 245$

3. Hipótesis alternativa: H_a: $\mu \neq 245$

4. Fórmula para el valor del estadístico de prueba: $z = \dfrac{\bar{x} - 245}{s/\sqrt{n}}$

5. Cálculo del valor del estadístico de prueba: $z = \dfrac{246.18 - 245}{3.60/\sqrt{50}} = 2.32$

6. Determinación del valor *P*: como la prueba es de dos colas,

$$\text{Valor } P = 2(1 - \Phi(2.32)) = .0204$$

7. Conclusión: con un nivel significativo de .01, H_0 no sería rechazada puesto que $.0204 > .01$. A este nivel de significancia, no existe suficiente evidencia para concluir que el espesor promedio verdadero difiere del valor objetivo. ■

Valores *P* para pruebas *t*

Así como el valor *P* para una prueba *z* es un área de curva *z*, el valor *P* para una prueba *t* será un área de curva *t*. La figura 8.10 en la siguiente página ilustra los tres casos diferentes. El número de grados de libertad para la prueba *t* con una muestra es $n - 1$.

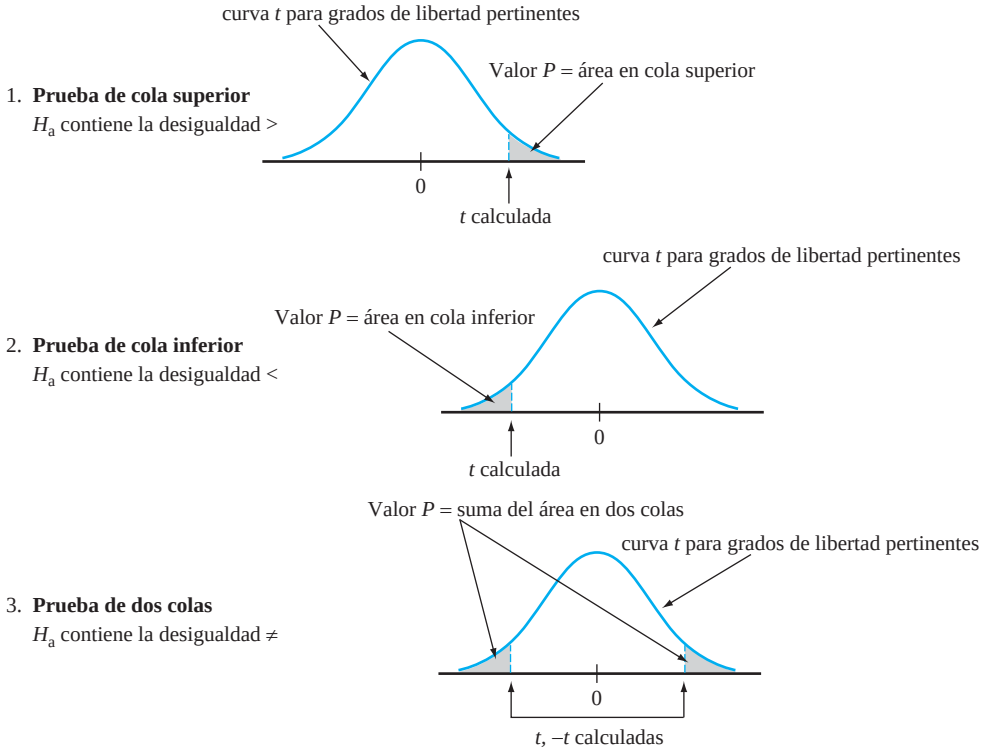


Figura 8.10 Valores *P* para pruebas *t*

La tabla de valores críticos *t* previamente utilizada para intervalos de confianza y predicción no contienen suficiente información sobre cualquier distribución *t* particular que permita la determinación precisa de áreas deseadas. Así que se ha incluido otra tabla *t* en la tabla A.8 del apéndice, una que contiene una tabulación de áreas de cola superior de curva *t*. Cada columna diferente de la tabla es para un número distinto de grados de libertad y las filas son para valores calculados del estadístico de prueba *t* que van desde 0.0 hasta 4.0 en incrementos de .1. Por ejemplo, el número .074 aparece en la intersección de la fila 1.6 y la columna de 8 grados de libertad, por lo que el área bajo la curva de 8 grados de libertad a la derecha de 1.6 (un área de cola superior) es .074. Como las curvas *t* son simétricas, .074 también es el área bajo la curva de 8 grados de libertad a la izquierda de -1.6 (un área de cola inferior).

Supóngase, por ejemplo, que una prueba de H_0: $\mu = 100$ contra H_a: $\mu > 100$ está basada en la distribución *t* de 8 grados de libertad. Si el valor calculado del estadístico de prueba es $t = 1.6$, entonces el valor *P* para esta prueba de cola superior es .074. Como .074 excede a .05, H_0 no podría ser rechazada a un nivel de significancia de .05. Si la hipótesis alternativa es H_a: $\mu < 100$ y una prueba basada en 20 grados de libertad da $t = -3.2$, entonces la tabla A.8 del apéndice muestra que el valor *P* es el área de cola inferior capturada .002. La hipótesis nula puede ser rechazada al nivel .05 o .01. Considérese probar H_0: $\mu_1 - \mu_2 = 0$ contra H_a: $\mu_1 - \mu_2 \neq 0$, la hipótesis nula afirma que las medias de las dos poblaciones son idénticas, en tanto que la hipótesis alternativa afirma que son diferentes sin especificar una dirección de alejamiento de H_0. Si una prueba *t* está basada en 20 grados de libertad y $t = 3.2$, entonces el valor *P* para esta prueba de dos colas es $2(.002) = .004$. Éste también sería el valor *P* para $t = -3.2$. El área de cola se duplica porque los valores tanto más grandes que 3.2 como más pequeños que -3.2 contradicen más a H_0 que lo que se calculó (valores alejados en *una u otra* colas de la curva *t*).

Ejemplo 8.18 En el ejemplo 8.9 se realizó una prueba de H_0: $\mu = 4$ contra H_a: $\mu \neq 4$ basada en muestra de $n = 5$ observaciones de una distribución normal de población. El estadístico de prueba calculado fue $-.594 \approx .6$. Si se examina la columna 4 $(= 5 - 1)$ grados de libertad de la tabla A.8 del apéndice hacia abajo hasta la fila .6, se ve que la entrada es .290. Debido a que se trata de una prueba de dos colas, esta área de la cola superior debe ser duplicada para obtener el valor *P* razonable. El resultado es el valor $P \approx .580$. Este valor *P* es claramente más grande que cualquier nivel de significancia α (.01, .05 e incluso .10), por lo que no hay razón para rechazar la hipótesis nula. Los datos de salida obtenidos con Minitab del ejemplo 8.9 incluyen un valor $P = .594$. Los valores *P* obtenidos con programas de computadora serán más precisos que los obtenidos de la tabla A.8 del apéndice puesto que los valores de *t* que aparecen en la tabla son precisos sólo a décimos de dígito. ■

Más sobre interpretación de los valores *P*

El valor *P* resultante de la realización de una prueba en una muestra seleccionada *no* es la probabilidad de que H_0 sea cierta, ni es la probabilidad de rechazar la hipótesis nula. Una vez más, es la probabilidad, calculada suponiendo que H_0 es cierta, de obtener un valor estadístico de la prueba por lo menos tan contradictorio para la hipótesis nula como el valor que realmente resultó. Por ejemplo, considere la prueba H_0: $\mu = 50$ en contra de H_0: $\mu < 50$ usando una prueba *z* de cola inferior. Si el valor calculado del estadístico de prueba es $z = -2.00$, entonces,

$$\begin{aligned}\text{valor } P &= P(Z < -2.00 \text{ cuando } \mu = 50)\\ &= \text{área bajo la curva } z \text{ a la izquierda de} -2.00 = 0.228\end{aligned}$$

Pero si se selecciona una segunda muestra, el valor resultante de *z* es casi seguro que será diferente de -2.00, por lo que el correspondiente valor *P* también es probable que difiera de .0228. Dado que el valor estadístico de la prueba varía por sí mismo de una muestra a otra, el valor *P* también puede variar de una muestra a otra. Es decir, el estadistico de prueba es una variable aleatoria, por lo que el valor *P* también será una variable aleatoria. Una primera muestra puede dar un valor *P* de .0228, una segunda muestra puede resultar en un valor *P* de .1175, una tercera puede producir .0606 como el valor *P* y así sucesivamente.

Si H_0 es falsa, se espera que el valor *P* sea cercano a 0 de manera que la hipótesis nula pueda ser rechazada. Por otro lado, cuando H_0 es cierta, sería bueno que el valor *P* excediera el nivel de significancia seleccionado de modo que se tome la decisión correcta al no rechazar H_0. El siguiente ejemplo presenta simulaciones para mostrar cómo se comporta el valor *P* cuando la hipótesis nula es verdadera y cuando es falsa.

Ejemplo 8.19 La eficiencia de combustible (mpg) de todos los vehículos nuevos particulares en determinadas condiciones de conducción no puede ser idéntica a la cifra de la EPA que aparece en la etiqueta del vehículo. Supóngase que cuatro vehículos diferentes de un tipo particular van a ser seleccionados y conducidos por un rumbo determinado, tras lo cual se determina el consumo de combustible de cada uno de ellos. Sea μ que denota la verdadera eficacia de combustible promedio en estas condiciones. Considérese la posibilidad de probar H_0: $\mu = 20$ contra H_0: $\mu > 20$ utilizando la prueba *t* de una muestra basada en la muestra resultante. Dado que la prueba se basa en $n - 1 = 3$ grados de libertad, el valor *P* para una prueba de cola superior es el área bajo la curva *t* con 3 gl a la derecha de la *t* calculada.

Primero se supondrá que la hipótesis nula es cierta. Se le pide a Minitab generar 10,000 muestras diferentes, cada una con 4 observaciones, a partir de una distribución normal de la población con valor medio $\mu = 20$ y desviación estándar $\sigma = 2$. La primera muestra y el resumen de las cantidades resultantes son

$$x_1 = 20.830,\ x_2 = 22.232,\ x_3 = 20.276,\ x_4 = 17.718$$

$$\bar{x} = 20.264 \quad s = 1.8864 \quad t = \frac{20.264 - 20}{.1.8864/\sqrt{4}} = .2799$$

El valor P es el área bajo la curva t de 3 gl a la derecha de .2799, que de acuerdo con Minitab es .3989. Al usar un nivel de significancia de .05, la hipótesis nula no sería, por supuesto, rechazada. Los valores de t para las siguientes cuatro muestras son -1.7591, .6082, $-.7020$ y 3.1053, con sus correspondientes valores P .912, .293, .733 y .0265.

La figura 8.11(a) muestra un histograma de los 10,000 valores P de este experimento de simulación. Cerca de 4.5% de estos valores P están en el intervalo de primera clase de 0 a .05. Así, cuando se utiliza un nivel de significancia de .05, la hipótesis nula se rechaza en aproximadamente 4.5% de estas 10,000 pruebas. Si se continúa para generar muestras y llevar a cabo la prueba para cada muestra a un nivel de significancia .05, en el largo plazo el 5% de los valores P estaría en el intervalo de primera clase. Esto es porque cuando H_0 es verdadera y la prueba se utiliza con un nivel de significancia .05, por definición, la probabilidad de rechazar H_0 es .05.

Observando el histograma, parece que la distribución de los valores P es relativamente plana. De hecho, se puede demostrar que, cuando H_0 es cierta, la distribución de probabilidad del valor P es una distribución uniforme en el intervalo de 0 a 1. Es decir, la curva de densidad es completamente plana en este intervalo y por lo tanto debe tener una altura de 1 si el área total bajo la curva es 1. Como el área bajo tal curva a la izquierda de .05 es (.05)(1) = .05, de nuevo tenemos que la probabilidad de rechazar H_0 cuando es cierta es .05, el nivel de significancia elegido.

Considérese ahora lo que sucede cuando H_0 es falsa ya que $\mu = 21$. Una vez más Minitab genera 10,000 muestras diferentes de tamaño 4 (cada una de una distribución normal con $\mu = 21$ y $\sigma = 2$), calcula $t = (\bar{x} - 20)/(s/\sqrt{4})$ para cada uno y luego determina el valor P. La primera de estas muestras resultó en $\bar{x} = 20.6411$, $s = .49637$, $t = 2.5832$, valor $P = .0408$. La figura 8.11(b) da un histograma de los valores P resultantes. La forma de este histograma es muy diferente de la de la figura 8.11(a), hay una tendencia mucho mayor para que el valor P sea pequeño (cercano a 0) cuando $\mu = 21$ que cuando $\mu = 20$. Una vez más se rechaza H_0 al nivel de significancia .05 cada vez que el valor P es cuando mucho .05 (en el intervalo de primera clase). Desafortunadamente, éste es el caso de sólo alrededor del 19% de los valores P. Así que sólo alrededor del 19% de las 10,000 pruebas rechazan correctamente la hipótesis nula; para el otro 81%, se comete un error de tipo II. La dificultad es que el tamaño de la muestra es muy pequeño y 21 no es muy diferente del valor declarado por la hipótesis nula.

La figura 8.11(c) ilustra lo que ocurre con el valor P cuando H_0 es falsa, porque $\mu = 22$ (todavía con $n = 4$ y $\sigma = 2$). El histograma está aún más concentrado hacia valores cercanos a 0 que en el caso cuando $\mu = 21$. En general, como μ se mueve más a la derecha del valor nulo 20, la distribución del valor P se hará más y más concentrada en valores cercanos a 0. Incluso aquí, un poco menos del 50% de los valores P son menores que .05. Por lo tanto, es todavía un poco más probable que la hipótesis nula no se rechace incorrectamente. Sólo para los valores de μ mucho mayores que 20 (por ejemplo, por lo menos 24 o 25) es muy probable que el valor P sea menor que .05 y dé así la conclusión correcta.

La idea principal de este ejemplo es que debido a que el valor de cualquier estadístico de prueba es aleatorio, el valor P también será una variable aleatoria y por lo tanto tiene una distribución. Cuanto más lejos esté el valor real del parámetro del valor especificado por la hipótesis nula, más se concentra la distribución del valor P en valores cercanos a 0 y mayor será la posibilidad de que la prueba rechace correctamente H_0 (que corresponde a una β más pequeña).

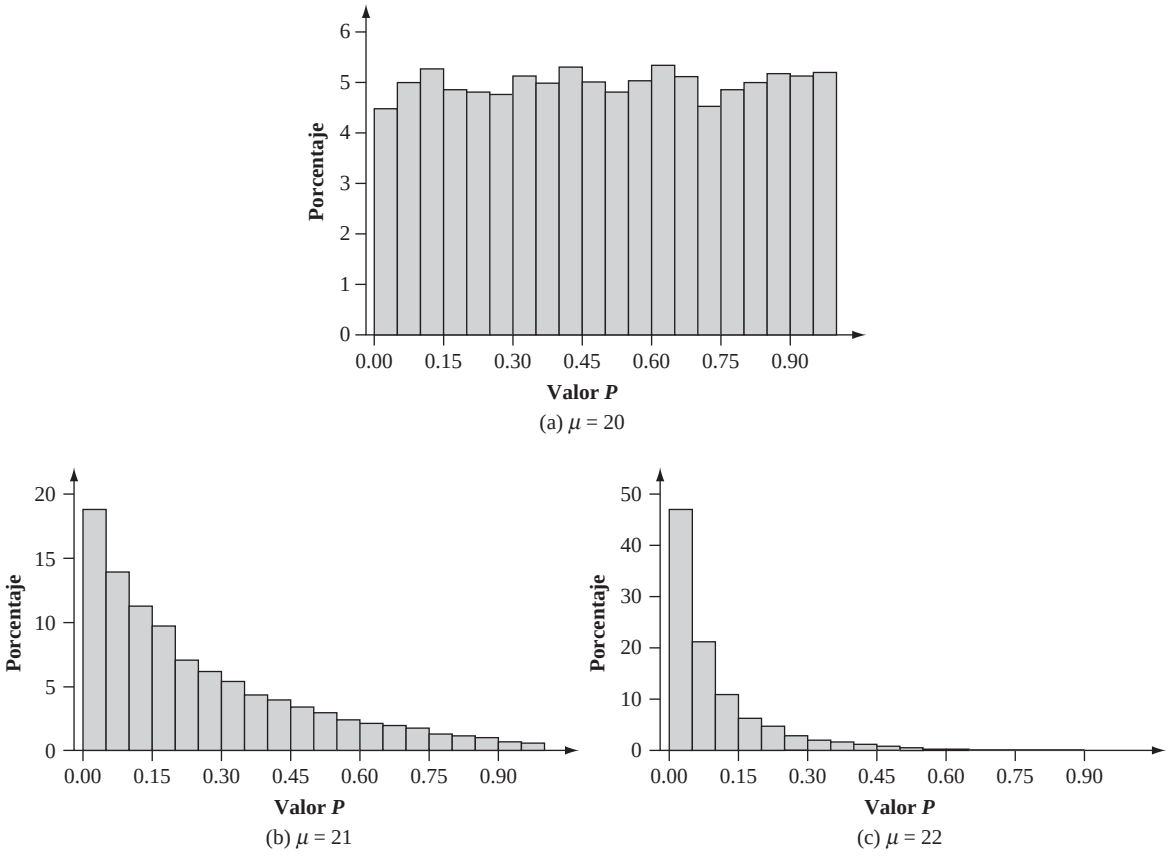


Figura 8.11 Valores *P* resultantes de la simulación para el ejemplo 8.19

EJERCICIOS Sección 8.4 (47–62)

47. ¿Con cuál de los valores *P* dados sería rechazada la hipótesis nula cuando se realiza una prueba de nivel .05?
a. .001 **b.** .021 **c.** .078
d. .047 **e.** .148

48. Se dan pares de valores *P* y niveles de significancia, α. Para cada par, diga si el valor *P* observado conduciría al rechazo de H_0 en el nivel de significancia dado.
a. Valor $P = .084$, $\alpha = .05$
b. Valor $P = .003$, $\alpha = .001$
c. Valor $P = .498$, $\alpha = .05$
d. Valor $P = .084$, $\alpha = .10$
e. Valor $P = .039$, $\alpha = .01$
f. Valor $P = .218$, $\alpha = .10$

49. Sea μ el tiempo medio de reacción a un estímulo. Para una prueba *z* con muestra grande de H_0: $\mu = 5$ contra H_a: $\mu > 5$, halle el valor *P* asociado con cada uno de los valores dados del estadístico de prueba *z*
a. 1.42 **b.** .90 **c.** 1.96 **d.** 2.48 **e.** −.11

50. Se supone que neumáticos de un tipo recién comprados están inflados a una presión de 30 lb/pulg^2. Sea μ la presión promedio verdadera. Halle el valor *P* asociado con cada valor estadístico *z* dado para probar H_0: $\mu = 30$ contra H_a: $\mu \neq 30$.
a. 2.10 **b.** −1.75 **c.** −.55 **d.** 1.41 **e.** −5.3

51. Dé tanta información como pueda sobre el valor *P* de una prueba *t* en cada una de las siguientes situaciones:
a. Prueba de cola superior, gl = 8, $t = 2.0$
b. Prueba de cola inferior, gl = 11, $t = -2.4$
c. Prueba de dos colas, gl = 15, $t = -1.6$
d. Prueba de cola superior, gl = 19, $t = -.4$
e. Prueba de cola superior, gl = 5, $t = 5.0$
f. Prueba de dos colas, gl = 40, $t = -4.8$

52. La pintura utilizada para trazar rayas en carreteras debe reflejar suficiente luz para que sea claramente visible de noche. Sea μ la lectura promedio verdadera del reflejómetro de un nuevo tipo de pintura considerada. Una prueba de H_0: $\mu = 20$ contra H_a: $\mu > 20$ se basará en una muestra aleatoria de tamaño *n* de una

distribución de población normal. ¿Qué conclusión es apropiada en cada una de las siguientes situaciones?

a. $n = 15$, $t = 3.2$, $\alpha = .05$
b. $n = 9$, $t = 1.8$, $\alpha = .01$
c. $n = 24$, $t = -.2$

53. Sea μ la concentración de receptor de suero en todas las mujeres embarazadas. Se sabe que el promedio de todas las mujeres es de 5.63. El artículo "Serum Transferrin Receptor for the Detection of Iron Deficiency in Pregnancy" (*Amer. J. of Clinical Nutr.*, 1991: 1077–1081) reporta ese valor $P > .10$ para una prueba de H_0: $\mu = 5.63$ contra H_a: $\mu \neq 5.63$ basada en $n = 176$ mujeres embarazadas. Con un nivel de significancia de .01, ¿qué concluiría?

54. El artículo "Analysis of Reserve and Regular Bottlings: Why Pay for a Difference Only the Critics Claim to Notice?" (*Chance*, verano de 2005, págs. 9–15) reportó sobre un experimento para investigar si los catadores de vino podían distinguir entre vinos de reserva más caros y sus contrapartes regulares. El vino fue presentado a los catadores en cuatro recipientes marcados A, B, C y D en dos de ellos con el vino de reserva y los otros dos con el vino regular. Cada catador seleccionó al azar tres de los recipientes, degustó los vinos seleccionados e indicó cuál de los tres creía que era diferente de los otros dos. De los $n = 855$ ensayos de degustación, 346 dieron por resultado la distinción correcta (el de reserva que difería de los dos vinos regulares o el vino regular que difería de dos reservas). ¿Proporciona esto evidencia contundente para concluir que los catadores de este tipo tienen capacidad para distinguir entre vinos de reserva y regulares? Formule y pruebe las hipótesis pertinentes con el método del valor *P*. ¿Se siente particularmente impresionado con la capacidad de los catadores de distinguir entre los dos tipos de vino?

55. Un fabricante de aspirina llena los frascos por peso en lugar de por conteo. Como cada frasco debe contener 100 tabletas, el peso promedio por tableta deberá ser de 5 gramos. Cada una de las 100 tabletas tomadas de un lote muy grande es pesada y el resultado es un peso promedio muestral por tableta de 4.87 gramos y una desviación estándar muestral de .35 gramo. ¿Proporciona esta información una fuerte evidencia para concluir que la compañía no está llenando sus frascos como lo anuncia? Pruebe las hipótesis apropiadas con $\alpha = .01$ calculando primero el valor *P* y luego comparándolo con el nivel de significancia especificado.

56. Debido a la variabilidad en el proceso de fabricación, el punto de cedencia de una muestra de acero suave sometida a un esfuerzo creciente normalmente diferirá del punto de cedencia teórico. Sea p la proporción verdadera de muestras que ceden antes de su punto de cedencia teórico. Si basándose en una muestra se concluye que más de 20% de todos los especímenes ceden antes del punto teórico, el proceso de producción tendrá que ser modificado.

a. Si 15 de 60 especímenes ceden antes del punto teórico, ¿cuál es el valor *P* cuando se utiliza la prueba apropiada y qué le aconsejaría hacer a la compañía?
b. Si el porcentaje verdadero de "cedencias tempranas" es en realidad de 50% (de modo que el punto teórico sea la mediana de la distribución de cedencia) y se utiliza una prueba de nivel .01, ¿cuál es la probabilidad de que la compañía concluya que es necesario modificar el proceso?

57. El artículo "Heavy Drinking and Polydrug Use Among College Students" (*J. of Drug Issues*, 2008: 445−466) señala que 51 de los 462 estudiantes universitarios en una muestra tenían una vida de abstinencia de alcohol. ¿Esto proporciona una fuerte evidencia para concluir que más del 10% de la población analizada se había abstenido por completo del consumo de alcohol? Pruebe la hipótesis apropiada utilizando el método del valor *P*. [*Nota:* el artículo utiliza los métodos estadísticos más avanzados para estudiar el uso de diversas drogas entre los estudiantes caracterizados como bebedores leves, moderados y fuertes.]

58. Se obtuvo una muestra aleatoria de especímenes de suelo y se determinó la cantidad (%) de materia orgánica presente en él por cada espécimen y se obtuvieron los datos adjuntos (tomados de "Engineering Properties of Soil", *Soil Science*, 1998: 93–102).

1.10	5.09	0.97	1.59	4.60	0.32	0.55	1.45
0.14	4.47	1.20	3.50	5.02	4.67	5.22	2.69
3.98	3.17	3.03	2.21	0.69	4.47	3.31	1.17
0.76	1.17	1.57	2.62	1.66	2.05		

Los valores de la media muestral, desviación estándar muestral y error estándar (estimado) de la media son 2.481, 1.616 y .295, respectivamente. ¿Sugieren estos datos que el porcentaje promedio verdadero de materia orgánica presente en el suelo es algún otro diferente de 3%? Realice una prueba de la hipótesis apropiada a un nivel de significancia de .10 determinando primero el valor *P*. ¿Sería diferente su conclusión si se hubiera usado $\alpha = .05$? [*Nota*: una gráfica de probabilidad normal de los datos muestra un patrón aceptable a la luz del tamaño de muestra razonablemente grande.]

59. Los datos que acompañan la fuerza del cubo de compresión (MPa) de probetas de hormigón apareció en el artículo "Experimental Study of Recycled Rubber-Filled High-Strength Concrete" (*Magazine of Concrete Res.*, 2009: 549−556):

112.3	97.0	92.7	86.0	102.0
99.2	95.8	103.5	89.0	86.7

a. ¿Es posible que la resistencia a la compresión para este tipo de concreto tenga una distribución normal?
b. Supóngase que el concreto se utilizará para una aplicación particular a menos que haya una fuerte evidencia de que la fuerza promedio real es inferior a 100 MPa. ¿Podría utilizarse el concreto? Realice una prueba de hipótesis adecuada utilizando el método del valor *P*.

60. Se diseñó una pluma de modo que el promedio verdadero de duración en condiciones controladas (implicando el uso de una máquina de escribir) sea por lo menos de 10 horas. Se seleccionó una muestra aleatoria de 18 plumas, se determinó la duración de cada una y una gráfica de probabilidad normal de los datos resultantes apoya el uso de una prueba *t* con una muestra.

a. ¿Qué hipótesis deberá ser probada si los investigadores creen *a priori* que la especificación de diseño ha sido satisfecha?
b. ¿Qué conclusión es apropiada si se prueban las hipótesis del inciso (a), $t = -2.3$ y $\alpha = .05$?

c. ¿Qué conclusión es apropiada si se prueban las hipótesis del inciso (a), $t = -1.8$ y $\alpha = .01$?

d. ¿Qué se deberá concluir si se prueban las hipótesis del inciso (a) y $t = -3.6$?

61. Un espectrofotómetro utilizado para medir concentración de CO [ppm (partes por millón) por volumen] se somete a prueba en cuanto a precisión tomando lecturas de un gas fabricado (llamado gas span) en el cual la concentración de CO se controla con precisión a 70 ppm. Si las lecturas sugieren que el espectrofotómetro no está funcionando de manera apropiada, éste tendrá que ser realizado. Suponga que es adecuadamente calibrado y la concentración medida en muestras de gas span está normalmente distribuida. Con base en las seis lecturas: 85, 77, 82, 68, 72 y 69, ¿es necesaria una recalibración? Realice una prueba de las hipótesis pertinentes utilizando el método del valor P con $\alpha = .05$.

62. La conductividad relativa de un dispositivo semiconductor está determinado por la cantidad de impurezas "adicionadas" al dispositivo durante su fabricación. Un diodo de silicio usado para propósitos específicos requiere un voltaje de corte promedio de .60 V y si éste no se alcanza, la cantidad de impurezas debe ser ajustada. Se seleccionó una muestra de diodos y se determinó el voltaje de corte. Los datos de salida adjuntos obtenidos con SAS son el resultado de una solicitud para probar las hipótesis apropiadas.

```
N        Mean     Std Dev          T  Prob. > |T|
15  0.0453333  0.0899100  1.9527887       0.0711
```

[*Nota*: SAS prueba explícitamente H_0: $\mu = 0$, así que para probar H_0: $\mu = .60$, el valor nulo .60 debe ser restado de cada x_i; la media reportada es entonces el promedio de los valores $(x_i - .60)$. También, el valor P de SAS siempre es para una prueba de dos colas.] ¿Qué se concluiría con un nivel de significancia de .01?, ¿de .05?, ¿de .10?

8.5 Algunos comentarios sobre la selección de una prueba

Una vez que el experimentador ha decidido sobre la cuestión de interés y el método de obtención de datos (el diseño del experimento), la construcción de una prueba apropiada se compone de tres pasos distintos:

1. Especificar un estadístico de prueba (la función de los valores observados que servirá para tomar una decisión).
2. Decidir sobre la forma general de la región de rechazo (típicamente rechazar H_0 con valores apropiadamente grandes del estadístico de prueba, rechazar con valores apropiadamente pequeños o rechazar con valores pequeños o grandes).
3. Seleccionar el valor o valores críticos numéricos específicos que separarán la región de rechazo de la región de aceptación (obteniendo la distribución del estadístico de prueba cuando H_0 es verdadera y luego seleccionar un nivel de significancia).

En los ejemplos presentados hasta ahora, se realizaron los pasos 1 y 2 en una manera adecuada mediante intuición. Por ejemplo, cuando la población subyacente se supuso normal con media μ y σ conocida, se procedió desde $\overline{X}$ hasta el estadístico de prueba estandarizado

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$$

Para probar H_0: $\mu = \mu_0$ contra H_a: $\mu > \mu_0$, la intuición sugirió entonces rechazar H_0 cuando z era grande. Por último, se determinó el valor crítico especificando el nivel de significancia α y utilizando el hecho de que Z tiene una distribución normal estándar cuando H_0 es verdadera. La confiabilidad de la prueba para tomar la decisión correcta puede ser evaluada estudiando probabilidades de error de tipo II.

Los temas que tienen que ser considerados al realizar los pasos 1–3 comprenden las preguntas:

1. ¿Cuáles son las implicaciones y consecuencias prácticas de seleccionar un nivel de significancia particular una vez que se han determinado los demás aspectos de una prueba?
2. ¿Existe un principio general, que no dependa sólo de la intuición, que pueda ser utilizado para obtener buenos o mejores procedimientos de prueba?

3. Cuando dos o más pruebas son apropiadas en una situación dada, ¿cómo se comparan las pruebas para decidir cuál deberá ser utilizada?
4. Si una prueba se realiza con arreglo a suposiciones específicas sobre la distribución o población muestreada, ¿cómo funcionará la prueba cuando se violan las suposiciones?

Significancia estadística contra práctica

Aunque el proceso de llegar a una decisión utilizando la metodología de probar hipótesis clásicas implica seleccionar un nivel de significancia y luego rechazar o no rechazar H_0, a ese nivel α, reportando simplemente el α utilizado y la decisión alcanzada conlleva poca de la información contenida en los datos muestrales. En especial, cuando los resultados de un experimento han de ser comunicados a una gran audiencia, el rechazo de H_0 a nivel de .05 será mucho más convincente si el valor observado del estadístico de prueba excede en gran medida el valor crítico de 5% que si apenas excede ese valor. Esto es precisamente lo que condujo a la noción de valor P como una forma de reportar significancia sin imponer un α particular a otros que pudieran desear sacar sus propias conclusiones.

Incluso si se incluye un valor P en un resumen de resultados, sin embargo, puede haber dificultad al interpretar este valor y al tomar una decisión. Esto es porque un valor P pequeño, el que ordinariamente indicaría **significancia estadística** en que sugeriría con fuerza el rechazo de H_0 a favor de H_a, puede ser el resultado de un tamaño de muestra grande en combinación con un alejamiento de H_0 que tiene poca **significancia práctica.** En muchas situaciones experimentales, sólo valdría la pena detectar los alejamientos de H_0 de gran magnitud, en tanto que un alejamiento pequeño de H_0 tendría poca significancia práctica.

Considérese como ejemplo probar H_0: $\mu = 100$ contra H_a: $\mu > 100$ donde μ es la media de una población normal con $\sigma = 10$. Supóngase que un valor verdadero de $\mu = 101$ no representaría un alejamiento serio de H_0 en el sentido de que no rechazar H_0 cuando $\mu = 101$ sería un error relativamente barato. Con tamaño de muestra razonablemente grande n, esta μ conduciría a un valor $\bar{x}$ próximo a 101, así que no se desearía esta evidencia muestral para argumentar fuertemente a favor del rechazo de H_0 cuando $\bar{x} = 101$ es observado. Para varios tamaños muestrales, la tabla 8.1 registra el valor P cuando $\bar{x} = 101$ y también la probabilidad de no rechazar H_0 al nivel .01 cuando $\mu = 101$.

La segunda columna en la tabla 8.1 muestra que incluso con tamaños de muestra moderadamente grandes, el valor P de $\bar{x} = 101$ argumenta con fuerza a favor del rechazo de H_0, en tanto que el valor $\bar{x}$ observado por sí mismo sugiere que en términos prácticos el valor verdadero de μ difiere poco del valor nulo $\mu_0 = 100$. La tercera columna señala que incluso cuando existe poca diferencia práctica entre la μ verdadera y el valor nulo, con un nivel de significancia fijo un tamaño de muestra grande casi siempre conduce al rechazo de la hipótesis nula a ese nivel. Resumiendo, *se debe tener un especial cuidado al interpretar evidencia cuando el tamaño de muestra es grande, puesto que cualquier alejamiento pequeño de* H_0 *casi con seguridad será detectado por una prueba, aunque semejante alejamiento pueda tener poca significancia práctica*.

Tabla 8.1 Ilustración del efecto del tamaño de muestra en los valores *P* y β

n	Valor P cuando $\bar{x} = 101$	$\beta(101)$ prueba de nivel .01
25	.3085	.9664
100	.1587	.9082
400	.0228	.6293
900	.0013	.2514
1600	.0000335	.0475
2500	.000000297	.0038
10,000	7.69×10^{-24}	.0000

El principio de razón de probabilidad

Sean $x_1, x_2, \ldots, x_n$ las observaciones en una muestra aleatoria de tamaño n de una distribución de probabilidad $f(x; \theta)$. La distribución conjunta evaluada con estos valores muestrales es el producto $f(x_1; \theta) \cdot f(x_2; \theta) \cdot \cdots \cdot f(x_n; \theta)$. Como en la discusión de estimación de probabilidad máxima, la *función de probabilidad* es esta distribución conjunta considerada como una función de θ. Considérese probar H_0: θ está en Ω_0 contra H_a: θ está en Ω_a, donde Ω_0 y Ω_a están desarticuladas (por ejemplo, H_0: $\theta \leq 100$ contra H_a: $\theta > 100$). El **principio de razón de probabilidad** para la construcción de una prueba prosigue como sigue:

1. Determinar el valor más grande de la probabilidad de que cualquier θ en Ω_0 (determinando la estimación de probabilidad máxima dentro de Ω_0 y sustituyendo de vuelta en la función de probabilidad).
2. Determinar el valor más grande de la probabilidad para cualquier θ en Ω_a.
3. Formar la razón

$$\lambda(x_1, \ldots, x_n) = \frac{\text{máxima probabilidad de } \theta \text{ en } \Omega_0}{\text{máxima probabilidad de } \theta \text{ en } \Omega_a}$$

La razón $\lambda(x_1, \ldots, x_n)$ se llama *valor estadístico de razón de probabilidad*. El procedimiento de prueba consiste en rechazar H_0 cuando esta razón es pequeña. Es decir, se elige una constante k y H_0 es rechazada si $\lambda(x_1, \ldots, x_n) \leq k$. Así pues H_0 es rechazada cuando el denominador de λ excede en gran medida al numerador, lo que indica que los datos son mucho más compatibles con H_a que con H_0.

La constante k se selecciona para que dé la probabilidad de error de tipo I deseada. Con frecuencia la desigualdad $\lambda \leq k$ puede ser manipulada para que produzca una condición equivalente más simple. Por ejemplo, para probar H_0: $\mu \leq \mu_0$ contra H_a: $\mu > \mu_0$ en el caso de normalidad, $\lambda \leq k$ equivale a $t \geq c$. Por consiguiente, con $c = t_{\alpha,n-1}$, la prueba de razón de probabilidad es la prueba t con una muestra.

El principio de razón de probabilidad también se aplica cuando las X_i tienen diferentes distribuciones e incluso cuando son dependientes, aunque la función de probabilidad puede ser complicada en tales casos. Muchos de los procedimientos de prueba que se presentarán en capítulos subsiguientes se obtienen a partir del principio de razón de probabilidad. Estas pruebas a menudo reducen al mínimo β entre todas las pruebas que tienen el nivel α deseado, así que verdaderamente son pruebas mejores. Para más detalles y algunos ejemplos resueltos, remítase a una de las referencias que aparecen en la bibliografía del capítulo 6.

Una limitación práctica para el uso del principio de razón de probabilidad es que, para construir el estadístico de prueba de razón de probabilidad, la forma de la distribución de probabilidad de donde proviene la muestra debe ser especificada. Para obtener la prueba t a partir del principio de razón de probabilidad, el investigador debe suponer una función de densidad de probabilidad normal. Si un investigador desea suponer que la distribución es simétrica pero no desea que sea específica con respecto a su forma exacta (tal como normal, uniforme o de Cauchy), en ese caso el principio falla porque no existe una forma de escribir una función de densidad de probabilidad conjunta válida al mismo tiempo para todas las distribuciones simétricas. En el capítulo 15 se presentarán varios procedimientos de prueba **libres de distribución,** llamados así porque la probabilidad de un error de tipo I es controlada simultáneamente para muchas distribuciones subyacentes diferentes. Estos procedimientos son útiles cuando el investigador tiene un conocimiento limitado de la distribución subyacente. Se dirá más sobre las cuestiones 3 y 4 listadas al principio de esta sección.

EJERCICIOS Sección 8.5 (63–64)

63. Reconsidere el problema de secado de pintura discutido en el ejemplo 8.2. Las hipótesis fueron H_0: $\mu = 75$ contra H_a: $\mu < 75$, suponiendo que el valor de σ es 9.0. Considere el valor alternativo $\mu = 74$, que en el contexto del problema de manera presumible no sería un alejamiento prácticamente significativo de H_0.

a. Con una prueba de nivel .01, calcule β para esta alternativa con tamaños de muestra $n = 100$, 900 y 2500.
b. Si el valor observado de $\bar{X}$ es $\bar{x} = 74$, ¿qué puede decir sobre el valor P resultante cuando $n = 2500$? ¿Son los datos estadísticamente significativos con cualquiera de los valores estándar de α?
c. ¿Realmente preferiría utilizar un tamaño de muestra de 2500 junto con una prueba de nivel .01 (haciendo caso omiso del costo de semejante experimento)? Explique.

64. Considere la prueba de nivel .01 con muestra grande en la sección 8.3 para probar H_0: $p = .2$ contra H_a: $p > .2$.
a. Para el valor alternativo $p = .21$, calcule $\beta(.21)$ con tamaños de muestra $n = 100$, 2500, 10,000, 40,000 y 90,000.
b. Para $\hat{p} = x/n = .21$, calcule el valor P cuando $n = 100$, 2500, 10,000 y 40,000.
c. En la mayoría de las situaciones, ¿sería razonable utilizar una prueba de nivel .01 junto con un tamaño de muestra de 40,000? ¿Por qué sí o por qué no?

EJERCICIOS SUPLEMENTARIOS (65–87)

65. Una muestra de 50 lentes utilizados en anteojos da un espesor medio muestral de 3.05 mm y una desviación estándar muestral de .34 mm. El espesor promedio verdadero deseado de los lentes es de 3.20 mm. ¿Sugieren los datos fuertemente que el espesor promedio verdadero de los lentes es algún otro diferente del deseado? Haga la prueba con $\alpha = .05$.

66. En el ejercicio 65, suponga que el experimentador creía antes de recopilar los datos que el valor de σ era de manera aproximada de .30. Si el experimentador deseó que la probabilidad de un error de tipo II fuera de .05 cuando $\mu = 3.00$, ¿era innecesariamente grande un tamaño de muestra 50?

67. Se especificó que cierto tipo de hierro debía contener .85 g de silicio por cada 100 g de hierro (.85%). Se determinó el contenido de silicio de cada uno de 25 especímenes seleccionados al azar y se obtuvieron los siguientes resultados con Minitab a partir de una prueba de las hipótesis apropiadas.

```
Variable    N    Mean    StDev  SE Mean     T    P
sil cont   25  0.8880   0.1807   0.0361  1.05 0.30
```

a. ¿Qué hipótesis se probaron?
b. ¿A qué conclusión llegaría con un nivel de significancia de .05 y por qué? Responda la misma pregunta para un nivel de significancia de .10.

68. Un método de enderezar alambre antes de enrollarlo para fabricar resortes se llama "enderezado con rodillos". El artículo "The Effect of Roller and Spinner Wire Straightening on Coiling Performance and Wire Properties" (*Springs*, 1987: 27–28) reporta sobre las propiedades de la tensión del alambre. Suponga que se selecciona una muestra de 16 alambres y cada uno se somete a prueba para determinar su resistencia a la tensión (N/mm^2). La media y desviación estándar muestrales resultantes son 2160 y 30, respectivamente.
a. La resistencia media a la tensión de resortes hechos mediante una máquina enderezadora rotatoria es de 2150 N/mm^2. ¿Qué hipótesis deberán ser probadas para determinar si la resistencia media a la tensión del método de rodillos excede de 2150?
b. Suponiendo que la distribución de la resistencia a la tensión es aproximadamente normal, ¿qué estadístico de prueba utilizaría para probar la hipótesis del inciso (a)?
c. ¿Cuál es el valor del estadístico de prueba con estos datos?
d. ¿Cuál es el valor P con el valor del estadístico de prueba calculado en el inciso (c)?
e. Con una prueba de nivel .05, ¿a qué conclusión llegaría?

69. La contaminación de los suelos de minas en China es un grave problema ambiental. El artículo "Heavy Metal Contamination in Soils and Phytoaccumulation in a Manganese Mine Wasteland, South China" (*Air, Soil, and Water Res.*, 2008: 31−41) informó que, para una muestra de 3 especímenes de suelo de cierta área restaurada de minería, la media de la concentración total de la muestra de Cu fue 45.31 mg/kg, con un correspondiente error estándar (estimado) de la media de 5.26. Se dijo también que el valor histórico en China de esta concentración fue de 20. Los resultados de diversas pruebas estadísticas descritas en el artículo se basan en el supuesto de normalidad.
a. ¿Los datos proporcionan una fuerte evidencia para concluir que la concentración real promedio en la región de la muestra supera el valor histórico planteado? Lleve a cabo una prueba al nivel de significancia .01 utilizando el método de valor P. ¿Le sorprende el resultado? Explique.
b. Volviendo a la prueba del inciso (a), ¿qué tan probable es que el valor P fuera al menos .01 cuando la concentración promedio real es de 50 y la verdadera desviación estándar de la concentración es de 10?

70. El artículo "Orchard Floor Management Utilizing Soil Applied Coal Dust for Frost Protection" (*Agri. and Forest Meteorology*, 1988: 71–82) reporta los siguientes valores de flujo de calor a través del suelo de ocho solares cubiertos con polvo de hulla.

34.7 35.4 34.7 37.7 32.5 28.0 18.4 24.9

El flujo de calor medio a través del suelo en solares cubiertos sólo con césped es de 29.0. Suponiendo que la distribución del flujo de calor es aproximadamente normal, ¿sugieren los datos que el polvo de hulla es eficaz para incrementar el flujo medio de calor sobre el del césped? Pruebe las hipótesis apropiadas con $\alpha = .05$.

71. El artículo "Caffeine Knowledge, Attitudes, and Consumption in Adult Women" (*J. of Nutrition Educ.*, 1992: 179–184) reporta los siguientes datos sobre consumo diario de cafeína para una muestra de mujeres adultas: $n = 47$, $\bar{x} = 215$ mg, $s = 235$ mg y rango = 5–1176.
a. ¿Parece posible que la distribución de la población de consumo diario de cafeína sea normal? ¿Es necesario suponer una distribución de población normal para probar hipótesis acerca del valor del consumo medio de la población? Explique su razonamiento.

b. Suponga que previamente se creía que el consumo medio era cuando mucho de 200 mg. ¿Contradicen los datos dados esta creencia previa? Pruebe las hipótesis apropiadas a nivel de significancia de .10 e incluya un valor *P* en su análisis.

72. El volumen de negocios anual de un fondo de inversión es el porcentaje de los activos de un fondo que se venden en un año determinado. En términos generales, un fondo con un valor bajo de volumen de negocios es más estable y contrario al riesgo, mientras que un valor alto de volumen de negocios indica una cantidad sustancial de compra y venta en un intento de tomar ventaja de las fluctuaciones del mercado a corto plazo. Los siguientes son los valores del volumen de negocios para una muestra de 20 fondos de gran capitalización mezclados (véase el ejercicio 1.53 para un poco más de información) extraídos de Morningstar.com:

1.03 1.23 1.10 1.64 1.30 1.27 1.25 0.78 1.05 0.64
0.94 2.86 1.05 0.75 0.09 0.79 1.61 1.26 0.93 0.84

a. ¿Usaría la prueba *t* de una muestra para decidir si existen pruebas convincentes para concluir que la población media de volumen de negocios es inferior al 100%? Explique.

b. Una gráfica de probabilidad normal del 20 ln(el volumen de negocios) de los valores muestra un patrón lineal muy pronunciado, lo que sugiere que es razonable suponer que la distribución del volumen de negocios es lognormal. Recuérdese que X tiene una distribución logarítmica normal si $\ln(X)$ es una distribución normal con valor medio μ y varianza σ^2. Como μ también es la mediana de la distribución $\ln(X)$, e^{μ} es la mediana de la distribución X. Utilice esta información para decidir si existen pruebas convincentes para concluir que la mediana de la distribución de la población del volumen de negocios es inferior a 100%.

73. Se supone que la resistencia a la ruptura promedio verdadera de aislantes de cerámica de cierto tipo es por lo menos de 10 lb/pulg^2. Se utilizará para una aplicación particular a menos que los datos muestrales indiquen concluyentemente que esta especificación no ha sido satisfecha. Una prueba de hipótesis con $\alpha = .01$ tiene que basarse en una muestra aleatoria de diez aislantes. Suponga que la distribución de resistencia a la ruptura es normal con desviación estándar desconocida.

a. Si la desviación estándar verdadera es de .80, ¿qué tan probable es que los aislantes sean juzgados satisfactorios cuando la resistencia a la ruptura promedio verdadera es en realidad de sólo 9.5? ¿Sólo de 9.0?

b. ¿Qué tamaño de muestra sería necesario para tener 75% de posibilidad de detectar que la resistencia a la ruptura promedio verdadera es de 9.5 cuando la desviación estándar verdadera es de .80?

74. Las observaciones adjuntas sobre tiempo de permanencia de llamas (s) en tiras de ropa de dormir de niños tratada aparecieron en el artículo "An Introduction to Some Precision and Accuracy of Measurement Problems" (*J. of Testing and Eval*., 1982: 132–140). Suponga que se había asignado por encargo un tiempo de permanencia de llamas promedio verdadero de cuando mucho 9.75. ¿Sugieren los datos que esta condición no se ha cumplido? Realice una prueba apropiada después de investigar la plausibilidad de las suposiciones que fundamentan su método de inferencia.

9.85 9.93 9.75 9.77 9.67 9.87 9.67
9.94 9.85 9.75 9.83 9.92 9.74 9.99
9.88 9.95 9.95 9.93 9.92 9.89

75. Se cree que la incidencia de un tipo de cromosoma defectuoso en la población de varones adultos estadounidenses es de 1 en 75. Una muestra aleatoria de 800 individuos en instituciones penitenciarias estadounidenses revela que 16 tienen tales defectos. ¿Se puede concluir que la proporción de incidencia de este defecto entre los prisioneros difiere de la proporción supuesta para toda la población de varones adultos?

a. Formule y pruebe las hipótesis pertinentes con $\alpha = .05$. ¿Qué tipo de error podría haber cometido al llegar a una conclusión?

b. ¿Qué valor *P* está asociado con esta prueba? Basado en este valor *P*, ¿podría H_0 ser rechazada a un nivel de significancia de .20?

76. En una investigación de la toxina producida por una serpiente venenosa, un investigador preparó 26 frascos, cada uno con 1 g de la toxina y luego determinó la cantidad de antitoxina necesaria para neutralizar la toxina. Se encontró que la cantidad promedio muestral de antitoxina necesaria era de 1.89 mg y la desviación estándar muestral era de .42. Una investigación previa indicó que la cantidad neutralizante promedio verdadera fue de 1.75 mg/g de toxina. ¿Contradicen estos datos nuevos el valor sugerido por la investigación previa? Pruebe la hipótesis pertinente usando el método de valor *P*. ¿Depende la validez de su análisis de cualquier suposición sobre la distribución de la población de cantidad neutralizante? Explique.

77. La resistencia a la compresión no restringida promedio muestral de 45 especímenes de un tipo particular de ladrillos resultó ser de 3107 lb/pulg^2 y la desviación estándar muestral fue de 188. La distribución de la resistencia a la compresión no restringida puede ser un tanto asimétrica. ¿Indican los resultados fuertemente que la resistencia a la compresión no restringida promedio verdadera es menor que el valor de diseño de 3200? Haga la prueba con $\alpha = .001$.

78. El 30 de diciembre de 2009, el *New York Times* informó que en una encuesta de 948 adultos estadounidenses que dijeron que estaban por lo menos un poco interesados en el futbol universitario, 597 dijo que el actual Bowl Championship System debe ser sustituido por una eliminatoria similar a la utilizada en el baloncesto universitario. ¿Esto aporta pruebas convincentes para concluir que la mayoría de todos los individuos están a favor de la sustitución del B.C.S. con una eliminatoria? Pruebe la hipótesis apropiada utilizando el método de valor *P*.

79. Cuando $X_1, X_2, \ldots, X_n$ son variables de Poisson independientes, cada una con parámetro μ, y la n es grande, la media muestral $\overline{X}$ tiene aproximadamente una distribución normal con $\mu = E(\overline{X})$ y $V(\overline{X}) = \mu/n$. Esto implica que

$$Z = \frac{\overline{X} - \mu}{\sqrt{\mu/n}}$$

tiene aproximadamente una distribución normal estándar. Para probar H_0: $\mu = \mu_0$, se puede reemplazar μ con μ_0 en la ecuación para Z a fin de obtener un estadístico de prueba. Normalmente se prefiere este estadístico al estadístico de muestra grande con denominador $S/\sqrt{n}$ (cuando las X_i son de Poisson) porque está explícitamente hecho a la medida de la suposición de Poisson. Si el número de solicitudes de consultoría recibidas por cierto estadístico durante una semana de trabajo de 5 días tiene una distribución de Poisson y el número total de solicitudes de consultoría durante 36 semanas es de 160, ¿sugiere esto que el número promedio verdadero de solicitudes semanales excede de 4.0? Haga la prueba con $\alpha = .02$.

80. Un artículo en el ejemplar del 11 de noviembre de 2005 del *Tribune* de San Luis Obispo reportó que los investigadores que realizan compras aleatorias en tiendas Wal-Mart en California encontraron que los escáneres dan el precio equivocado 8.3% del tiempo. Suponga que esto se basó en 200 compras. El National Institute for Standards and Technology comenta que a la larga cuando mucho dos de 100 artículos deberían tener precios incorrectamente escaneados.

a. Elabore un procedimiento de prueba con un nivel de significancia de (aproximadamente) .05 y luego realice la prueba para decidir si la referencia del NIST no se cumple.

b. Con el procedimiento de prueba empleado en (a), ¿cuál es la probabilidad de decidir que la referencia del NIST ha sido satisfecha cuando en realidad la proporción de errores es de 5%?

81. Un fabricante de tinas calientes anuncia que con su equipo de calefacción se puede alcanzar una temperatura de 100°F en 15 minutos en forma aproximada. Se selecciona una muestra aleatoria de 42 tinas y se determina el tiempo necesario para alcanzar una temperatura de 100°F con cada tina. El tiempo promedio y la desviación estándar muestrales son de 16.5 y 2.2 min, respectivamente. ¿Siembran estos dados alguna duda sobre la afirmación de la compañía? Calcule el valor P y utilícelo para llegar a una conclusión al nivel .05.

82. El capítulo 7 presentó un intervalo de confianza para la varianza σ^2 de una distribución de población normal. El resultado clave allí fue que la variable aleatoria $\chi^2 = (n-1)S^2/\sigma^2$ tiene una distribución ji cuadrada con $n-1$ grados de libertad. Considere la hipótesis nula H_0: $\sigma^2 = \sigma_0^2$ (en forma equivalente, $\sigma = \sigma_0$). Entonces cuando H_0 es verdadera, el estadístico de prueba $\chi^2 = (n-1)S^2/\sigma_0^2$ tiene una distribución ji cuadrada con $n-1$ grados de libertad. Si la alternativa pertinente es H_a: $\sigma^2 > \sigma_0^2$, rechazar H_0 si $(n-1)s^2/\sigma_0^2 \geq \chi^2_{\alpha,\,n-1}$ da una prueba con nivel de significancia α. Para garantizar características razonablemente uniformes en una aplicación particular, se desea que la desviación estándar verdadera del punto de ablandamiento de cierto tipo de alquitrán de petróleo sea cuando mucho de .50°C. Se determinaron los puntos de ablandamiento de diez diferentes especímenes y se obtuvo una desviación estándar muestral de .58°C. ¿Contradice esto fuertemente la especificación de uniformidad? Pruebe las hipótesis apropiadas con $\alpha = .01$.

83. Remitiéndose al ejercicio 82, suponga que un investigador desea probar H_0: $\sigma^2 = .04$ contra H_a: $\sigma^2 < .04$ basado en una muestra de 21 observaciones. El valor calculado de $20s^2/.04$ es 8.58. Ponga límites en el valor P y luego llegue a una conclusión al nivel .01.

84. Cuando la distribución de la población es normal y n es grande, la desviación estándar muestral S tiene aproximadamente una distribución normal con $E(S) \approx \sigma$ y $V(S) \approx \sigma^2/(2n)$. Ya se sabe que en este caso, con cualquier n, $\overline{X}$ es normal con $E(\overline{X}) = \mu$ y $V(\overline{X}) = \sigma^2/n$.

a. Suponiendo que la distribución subyacente es normal, ¿cuál es un estimador aproximadamente insesgado del 99° percentil $\theta = \mu + 2.33\sigma$?

b. Cuando las X_i son normales, se puede demostrar que $\overline{X}$ y S son variables aleatorias independientes (una mide la ubicación mientras que la otra mide la dispersión). Use esto para calcular $V(\hat{\theta})$ y $\sigma_{\hat{\theta}}$ para el estimador $\hat{\theta}$ del inciso (a). ¿Cuál es el error estándar estimado $\hat{\sigma}_{\hat{\theta}}$?

c. Escriba un estadístico de prueba para probar H_0: $\theta = \theta_0$ que tiene aproximadamente una distribución estándar normal cuando H_0 es verdadera. Si el pH del suelo está normalmente distribuido en cierta región y 64 muestras de suelo dan $\bar{x} = 6.33$, $s = .16$, ¿proporciona esto una fuerte evidencia para concluir que cuando mucho 99% de todas las muestras posibles tendrían un pH de menos de 6.75? Pruebe con $\alpha = .01$.

85. Sean $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución exponencial con parámetro λ. En ese caso se puede demostrar que $2\lambda\sum X_i$ tiene una distribución ji cuadrada con $\nu = 2n$ (demuestre primero que $2\lambda X_i$ tiene distribución ji cuadrada con $\nu = 2$).

a. Use este hecho para obtener un estadístico de prueba y región de rechazo que juntos especifiquen una prueba a nivel para H_0: $\mu = \mu_0$ contra cada una de las tres alternativas comúnmente encontradas. [*Sugerencia*: $E(X_i) = \mu = 1/\lambda$, de modo que $\mu = \mu_0$ equivale a $\lambda = 1/\mu_0$.]

b. Suponga que se prueban diez componentes idénticos que tienen un tiempo exponencialmente distribuido hasta la falla. Los tiempos de falla resultantes son

95 16 11 3 42 71 225 64 87 123

Use el procedimiento de prueba del inciso (a) para decidir si los datos sugieren fuertemente que la vida útil promedio verdadera es menor que el valor antes afirmado de 75.

86. Suponga que la distribución de la población es normal con σ conocida. Sea γ de modo que $0 < \gamma < \alpha$. Para probar H_0: $\mu = \mu_0$ contra H_a: $\mu \neq \mu_0$, considere la prueba que rechaza H_0 si $z \geq z_\gamma$ o $z \leq -z_{\alpha-\gamma}$, donde el estadístico de prueba es $Z = (\overline{X} - \mu_0)/(\sigma/\sqrt{n})$.

a. Demuestre que P(error de tipo I) $= \alpha$.

b. Deduzca una expresión para $\beta(\mu')$. [*Sugerencia*: exprese la prueba en la forma "rechazar H_0 si $\bar{x} \geq c_1$ o $\leq c_2$".]

c. Sea $\Delta > 0$. ¿Con qué valores de γ (con respecto a α) será $\beta(\mu_0 + \Delta) < \beta(\mu_0 - \Delta)$?

87. Luego de un periodo de aprendizaje una organización realiza un examen que debe ser aprobado a fin de ser elegible para una membresía. Sea $p = P$(aprendiz seleccionado al azar aprueba el examen). La organización desea un examen que la mayoría mas no todos deberá ser capaz de aprobar, por lo que decide que $p = .90$ es deseable. Para un examen particular, las hipótesis pertinentes son H_0: $p = .90$ contra la alternativa H_a: $p \neq .90$. Suponga que diez personas hacen el examen y sea $X =$ el número que aprueba el examen.

a. ¿La región de cola inferior $\{0, 1, \ldots, 5\}$ especifica una prueba de nivel .01?

b. Demuestre que aun cuando H_a es bilateral, ninguna prueba de dos colas es una prueba de nivel .01.

c. Trace una gráfica de $\beta(p')$ como función de p' para esta prueba. ¿Es esto deseable?

Bibliografía

Véanse las bibliografías al final de los capítulos 6 y 7.

9 Inferencias basadas en dos muestras

INTRODUCCIÓN

Los capítulos 7 y 8 presentaron intervalos de confianza (IC) y procedimientos de prueba de hipótesis para una sola media μ, una sola proporción p y una sola varianza σ^2. En este capítulo se amplían estos métodos a situaciones que implican las medias, las proporciones y las varianzas de dos distribuciones de población diferentes. Por ejemplo, sea μ_1 la dureza Rockwell promedio verdadera de especímenes de acero térmicamente tratados y μ_2 la dureza promedio verdadera de especímenes laminados en frío. Entonces es posible que un investigador desee utilizar muestras de observaciones de dureza de cada tipo de acero como base para calcular una estimación de intervalo de $\mu_1 - \mu_2$, la diferencia entre las dos durezas promedio verdaderas. Como otro ejemplo, sea p_1 la proporción verdadera de celdas de níquel-cadmio producidas en las condiciones de producción actuales defectuosas a causa de cortos internos y sea p_2 la proporción verdadera de celdas con cortos internos producidas en condiciones de operación modificadas. Si el razonamiento en cuanto a las condiciones modificadas es reducir la proporción de celdas defectuosas, un ingeniero de calidad desearía utilizar información muestral para probar la hipótesis nula H_0: $p_1 - p_2 = 0$ (es decir, $p_1 = p_2$) contra la hipótesis alternativa H_a: $p_1 - p_2 > 0$ (es decir, $p_1 > p_2$).

9.1 Pruebas *z* e intervalos de confianza para una diferencia entre dos medias de población

Las inferencias discutidas en esta sección se refieren a una diferencia $\mu_1 - \mu_2$ entre las medias de dos distribuciones de población diferentes. Un investigador podría, por ejemplo, desear probar hipótesis con respecto a la diferencia entre resistencias a la ruptura promedio verdaderas de dos tipos distintos de cartón corrugado. Una de esas hipótesis formularía que $\mu_1 - \mu_2 = 0$, es decir, que $\mu_1 = \mu_2$. Alternativamente, puede ser apropiado estimar $\mu_1 - \mu_2$ calculando un intervalo de confianza de 95%. Tales inferencias están basadas en una muestra de observaciones de resistencia de cada tipo de cartón.

Suposiciones básicas

1. $X_1, X_2, \ldots, X_m$ es una muestra aleatoria de una distribución con media μ_1 y varianza σ_1^2.
2. $Y_1, Y_2, \ldots, Y_n$ es una muestra aleatoria de una distribución con media μ_2 y varianza σ_2^2.
3. Las muestras X y Y son independientes entre sí.

El uso de m para el número de observaciones en la primera muestra y n el número de observaciones en la segunda muestra permite que los dos tamaños de la muestra sean diferentes. A veces esto se debe a que es más difícil o caro muestrear una población que otra. En otras situaciones, inicialmente puede especificarse igual tamaño de muestra, pero por razones más allá del alcance del experimento, el tamaño de las muestras reales puede diferir. Por ejemplo, el resumen del artículo "A Randomized Controlled Trial Assessing the Effectiveness of Professional Oral Care by Dental Hygienists" (*Intl. J. of Dental Hygiene*, 2008: 63–67) señala que "cuarenta pacientes fueron asignados al azar al grupo de POC ($m = 20$) o el grupo control ($n = 20$). Un paciente del grupo de POC y tres en el grupo control se retiraron a causa de la exacerbación de la enfermedad subyacente o la muerte". El análisis de los datos se basó entonces en $m = 19$ y $n = 16$.

El estimador natural de $\mu_1 - \mu_2$ es $\overline{X} - \overline{Y}$, la diferencia entre las medias muestrales correspondientes. Procedimientos inferenciales se basan en estandarizar este estimador, así que se requieren expresiones para el valor esperado y la desviación estándar de $\overline{X} - \overline{Y}$.

PROPOSICIÓN

El valor esperado de $\overline{X} - \overline{Y}$ es $\mu_1 - \mu_2$, así que $\overline{X} - \overline{Y}$ es un estimador insesgado de $\mu_1 - \mu_2$. La desviación estándar de $\overline{X} - \overline{Y}$ es

$$\sigma_{\overline{X}-\overline{Y}} = \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}$$

Demostración Estos dos resultados dependen de las reglas de valor esperado y varianza presentados en el capítulo 5. Como el valor esperado de una diferencia es la diferencia de valores esperados,

$$E(\overline{X} - \overline{Y}) = E(\overline{X}) - E(\overline{Y}) = \mu_1 - \mu_2$$

Como las muestras X y Y son independientes, $\overline{X}$ y $\overline{Y}$ son cantidades independientes, así que la varianza de la diferencia es la *suma* de $V(\overline{X})$ y $V(\overline{Y})$:

$$V(\overline{X} - \overline{Y}) = V(\overline{X}) + V(\overline{Y}) = \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}$$

La desviación estándar de $\overline{X} - \overline{Y}$ es la raíz cuadrada de esta expresión. ■

Si se piensa en $\mu_1 - \mu_2$ como un parámetro θ, entonces su estimador es $\hat{\theta} = \overline{X} - \overline{Y}$ con desviación estándar $\sigma_{\hat{\theta}}$ dada por la proposición. Cuando tanto σ_1^2 como σ_2^2 tienen valores conocidos, el valor de esta desviación estándar puede ser calculado. Las varianzas muestrales deben ser utilizadas para estimar $\sigma_{\hat{\theta}}$ cuando σ_1^2 y σ_2^2 son desconocidas.

Procedimientos de prueba para poblaciones normales con varianzas conocidas

En los capítulos 7 y 8, el primer intervalo de confianza y procedimiento de prueba para una media de población μ se basaron en la suposición de que la distribución de la población era normal con el valor de la varianza de población σ^2 conocido por el investigador. Asimismo, primero se supone que ambas distribuciones de población son normales y que los valores *tanto* de σ_1^2 como de σ_2^2 son conocidos. En breve se presentarán situaciones en las cuales una o ambas suposiciones pueden ser eximidas.

Como las distribuciones de población son normales, tanto $\overline{X}$ como $\overline{Y}$ tienen distribuciones normales. Por otra parte, la independencia de las dos muestras implica que las dos medias de las muestras son independientes una de otra. Esto implica que $\overline{X} - \overline{Y}$ está normalmente distribuida con valor esperado $\mu_1 - \mu_2$ y desviación estándar $\sigma_{\overline{X}-\overline{Y}}$ dada en la proposición precedente. Al estandarizar $\overline{X} - \overline{Y}$ se obtiene la variable normal estándar

$$Z = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \tag{9.1}$$

En un problema de prueba de hipótesis, la hipótesis nula formulará que $\mu_1 - \mu_2$ tiene un valor específico. Si Δ_0 denota este valor nulo, se tiene H_0: $\mu_1 - \mu_2 = \Delta_0$. Con frecuencia $\Delta_0 = 0$, en cuyo caso H_0 dice que $\mu_1 = \mu_2$. Al reemplazar $\mu_1 - \mu_2$ en la expresión (9.1) con el valor nulo Δ_0 se obtiene un estadístico de prueba. El estadístico de prueba Z se obtiene estandarizando $\overline{X} - \overline{Y}$ de conformidad con la suposición de que H_0 es verdadera, así que en este caso tiene una distribución estándar normal. Esta prueba estadística se puede escribir como $(\hat{\theta}$ – valor nulo$)/\sigma_{\hat{\theta}}$, que tiene la misma forma que varias de las estadísticas de prueba en el capítulo 8.

Considérese la hipótesis alternativa H_a: $\mu_1 - \mu_2 > \Delta_0$. Un valor $\bar{x} - \bar{y}$ que exceda considerablemente de Δ_0 (el valor esperado de $\overline{X} - \overline{Y}$ cuando H_0 es verdadera) proporciona evidencia en contra de H_0 y a favor de H_a. Tal valor de $\bar{x} - \bar{y}$ corresponde a un valor positivo y grande de z. Por consiguiente H_0 deberá ser rechazada a favor de H_a si z es mayor que o igual a un valor crítico apropiadamente seleccionado. Como el estadístico de prueba Z tiene una distribución normal estándar cuando H_0 es verdadera, la región de rechazo de cola superior $z \geq z_\alpha$ produce una prueba con nivel de significación (probabilidad de error de tipo I) α. Las regiones de rechazo para H_a: $\mu_1 - \mu_2 < \Delta_0$ y H_a: $\mu_1 - \mu_2 \neq \Delta_0$ que producen pruebas con nivel de significancia deseado α son la de cola inferior y la de dos colas, respectivamente.

Hipótesis nula: H_0: $\mu_1 - \mu_2 = \Delta_0$

Valor estadístico de prueba: $z = \dfrac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\dfrac{\sigma_1^2}{m} + \dfrac{\sigma_2^2}{n}}}$

Hipótesis alternativa	Región de rechazo para prueba con nivel α
H_a: $\mu_1 - \mu_2 > \Delta_0$	$z \geq z_\alpha$ (cola superior)
H_a: $\mu_1 - \mu_2 < \Delta_0$	$z \leq -z_\alpha$ (cola inferior)
H_a: $\mu_1 - \mu_2 \neq \Delta_0$	$z \geq z_{\alpha/2}$ o $z \leq -z_{\alpha/2}$ (dos colas)

Como éstas son pruebas z, se calcula un valor P como se hizo para las pruebas z en el capítulo 8 [p. ej., valor $P = 1 - \Phi(z)$ para una prueba de cola superior].

Ejemplo 9.1 Un análisis de una muestra aleatoria compuesta de $m = 20$ especímenes de acero laminado en frío para determinar las resistencias a ceder dio por resultado una resistencia promedio muestral de $\bar{x} = 29.8$ kg/pulg^2. Una segunda muestra aleatoria de $n = 25$ especímenes de acero galvanizado bilaterales dio una resistencia promedio muestral de $\bar{y} = 34.7$ kg/pulg^2. Suponiendo que las dos distribuciones de resistencia a ceder son normales con $\sigma_1 = 4.0$ y $\sigma_2 = 5.0$ (sugeridas por una gráfica en el artículo "Zinc-Coated Sheet Steel: An Overview", *Automotive Engr.*, diciembre de 1984: 39–43), ¿indican los datos que las resistencias a ceder promedio verdaderas μ_1 y μ_2 son diferentes? Realice una prueba a un nivel de significación $\alpha = .01$.

1. El parámetro de interés es $\mu_1 - \mu_2$, la diferencia entre las resistencias promedio verdaderas de los dos tipos de acero.
2. La hipótesis nula es H_0: $\mu_1 - \mu_2 = 0$.
3. La hipótesis alternativa es H_a: $\mu_1 - \mu_2 \neq 0$; si H_a es verdadera, entonces μ_1 y μ_2 son diferentes.
4. Con $\Delta_0 = 0$, el valor estadístico de prueba es
$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\dfrac{\sigma_1^2}{m} + \dfrac{\sigma_2^2}{n}}}$$
5. La desigualdad en H_a implica que la prueba es de dos colas. Con $\alpha = .01$, $\alpha/2 = .005$ y $z_{\alpha/2} = z_{.005} = 2.58$, H_0 será rechazada si $z \geq 2.58$ o si $z \leq -2.58$.
6. Sustituyendo $m = 20$, $\bar{x} = 29.8$, $\sigma_1^2 = 16.0$, $n = 25$, $\bar{y} = 34.7$ y $\sigma_2^2 = 25.0$ en la fórmula para z se obtiene
$$z = \frac{29.8 - 34.7}{\sqrt{\dfrac{16.0}{20} + \dfrac{25.0}{25}}} = \frac{-4.90}{1.34} = -3.66$$
Es decir, el valor observado de $\bar{x} - \bar{y}$ está a más de tres desviaciones estándar por debajo de lo que era de esperarse si H_0 fuera verdadera.
7. Como $-3.66 < -2.58$, z no queda en la cola inferior de la región de rechazo. H_0 es por consiguiente rechazada al nivel .01 a favor de la conclusión de que $\mu_1 \neq \mu_2$. Los datos muestrales sugieren fuertemente que la resistencia a ceder promedio verdadera del acero laminado en frío difiere de la del acero galvanizado. El valor P con esta prueba de dos colas es $2(1 - \Phi(3.66)) \approx 2(1 - 1) = 0$, de modo que H_0 debe ser rechazada a *cualquier* nivel de significancia razonable. ■

Utilización de una comparación para identificar causalidad

A los investigadores a menudo les interesa comparar los efectos de dos tratamientos diferentes en una respuesta o la respuesta después de un tratamiento con la respuesta de sin tratamiento (tratamiento vs. control). Si los individuos u objetos que van a ser utilizados en la comparación no son asignados por los investigadores a las dos diferentes condiciones, se dice que el estudio es **observacional.** La dificultad de sacar conclusiones basadas en un estudio observacional es que aunque el análisis estadístico puede indicar una diferencia significativa de respuesta entre los dos grupos, la diferencia puede deberse a algunos factores subyacentes que no habían sido controlados en lugar de a cualquier diferencia en los tratamientos.

Ejemplo 9.2 Una carta que apareció en el *Journal of the American Medical Association* (19 de mayo de 1978) reporta que de 215 médicos que se graduaron en Harvard y murieron entre noviembre de 1974 y octubre de 1977, 125 en servicio de tiempo completo vivieron un promedio de 48.9 años después de su graduación, en tanto que 90 con afiliaciones académicas vivieron un promedio de 43.2 años después de su graduación. ¿Sugieren estos datos que la vida media después de la graduación de doctores en práctica completa excede la vida media de aquellos que tienen afiliación académica? (De ser así, aquellos estudiantes que se "mueren por obtener una afiliación académica" pueden estar más cerca de la verdad de lo que realmente piensan, en otras palabras, ¿es "publicar o perecer" realmente "publicar o perecer"?)

Sea μ_1 el número promedio verdadero de años vividos después de la graduación de médicos en práctica completa, y μ_2 la misma cantidad de médicos con afiliaciones académicas. Suponga que los 125 y los 90 médicos son muestras aleatorias de las poblaciones 1 y 2, respectivamente (lo cual puede no ser razonable si hay razón para creer que los graduados de Harvard poseen características especiales que los diferencía de todos los demás médicos, en este caso las inferencias se limitarían sólo a las "poblaciones Harvard"). La carta de donde se tomaron los datos no dio información sobre varianzas, de modo que como ilustración supóngase que $\sigma_1 = 14.6$ y $\sigma_2 = 14.4$. Las hipótesis son H_0: $\mu_1 - \mu_2 = 0$ contra H_a: $\mu_1 - \mu_2 > 0$, de modo que Δ_0 es cero. El valor calculado del estadístico de prueba es

$$z = \frac{48.9 - 43.2}{\sqrt{\dfrac{(14.6)^2}{125} + \dfrac{(14.4)^2}{90}}} = \frac{5.70}{\sqrt{1.70 + 2.30}} = 2.85$$

El valor P para una prueba de cola superior es $1 - \Phi(2.85) = .0022$. A un nivel de significancia de .01, H_0 es rechazada (porque $\alpha >$ valor P) a favor de la conclusión de que $\mu_1 - \mu_2 > 0$ ($\mu_1 > \mu_2$). Esto es compatible con la información reportada en la carta.

Estos datos se derivaron de un estudio observacional **retrospectivo**; el investigador no comenzó seleccionando una muestra de doctores y asignando algunos al tratamiento de "afiliación académica" y a los demás al tratamiento de "práctica de tiempo completo", sino que en lugar de ello identificó miembros de los dos grupos reflexionando (¡mediante obituarios!) hasta observando registros pasados. ¿Puede ser el resultado estadísticamente significativo atribuido en realidad a una diferencia en el tipo de práctica médica después de la graduación, o existe algún otro factor subyacente (p. ej., edad al momento de la graduación, regímenes de ejercicio, etc.) que pudiera proporcionar también una explicación factible para la diferencia? Se han utilizado estudios observacionales para argumentar en cuanto a un vínculo causal entre el tabaquismo y el cáncer de pulmón. Existen muchos estudios que demuestran que la incidencia de cáncer de pulmón es significativamente más alta entre fumadores que entre no fumadores. No obstante, los individuos habían decidido si convertirse en fumadores mucho antes de que los investigadores aparecieran en la escena y los factores para tomar esta decisión pueden haber desempeñado un rol causal en la aparición de cáncer de pulmón. ■

Cuando investigadores asignan sujetos a los dos tratamientos de una manera aleatoria se obtiene un **experimento controlado aleatorizado.** Cuando se observa significancia estadística en semejante experimento, el investigador y otras partes interesadas tendrán más confianza en la conclusión de que la diferencia en la respuesta ha sido provocada por una diferencia en los tratamientos. Un ejemplo muy famoso de este tipo de experimento y conclusión es el experimento de la vacuna Salk contra la polio descrito en la sección 9.4. Estos temas son discutidos con mayor amplitud en los libros (no matemáticos) de Moore y de Freedman y colaboradores, incluidos en las referencias del capítulo 1.

β y la opción de tamaño de muestra

La probabilidad de un error de tipo II es fácil de calcular cuando ambas distribuciones de población son normales con valores conocidos de σ_1 y σ_2. Considérese el caso en el cual la hipótesis alternativa es H_a: $\mu_1 - \mu_2 > \Delta_0$. Sea Δ' un valor de $\mu_1 - \mu_2$ que excede Δ_0 (un valor con el cual H_0 es falsa). La región de rechazo de cola superior $z \geq z_\alpha$ puede ser reexpresada en la forma $\bar{x} - \bar{y} \geq \Delta_0 + z_\alpha \sigma_{\bar{X}-\bar{Y}}$. Por consiguiente

$$\begin{aligned}\beta(\Delta') &= P(\text{no rechazar } H_0 \text{ cuando } \mu_1 - \mu_2 = \Delta') \\ &= P(\bar{X} - \bar{Y} < \Delta_0 + z_\alpha \sigma_{\bar{X}-\bar{Y}} \text{ cuando } \mu_1 - \mu_2 = \Delta')\end{aligned}$$

Cuando $\mu_1 - \mu_2 = \Delta'$, $\bar{X} - \bar{Y}$ está normalmente distribuida con valor medio Δ' y desviación estándar $\sigma_{\bar{X}-\bar{Y}}$ (la misma desviación estándar como cuando H_0 es verdadera); con estos valores para estandarizar la desigualdad entre paréntesis da la probabilidad deseada.

Hipótesis alternativa	$\beta(\Delta') = P$(error de tipo II cuando $\mu_1 - \mu_2 = \Delta'$)
H_a: $\mu_1 - \mu_2 > \Delta_0$	$\Phi\left(z_\alpha - \dfrac{\Delta' - \Delta_0}{\sigma}\right)$
H_a: $\mu_1 - \mu_2 < \Delta_0$	$1 - \Phi\left(-z_\alpha - \dfrac{\Delta' - \Delta_0}{\sigma}\right)$
H_a: $\mu_1 - \mu_2 \neq \Delta_0$	$\Phi\left(z_{\alpha/2} - \dfrac{\Delta' - \Delta_0}{\sigma}\right) - \Phi\left(-z_{\alpha/2} - \dfrac{\Delta' - \Delta_0}{\sigma}\right)$

donde $\sigma = \sigma_{\bar{X}-\bar{Y}} = \sqrt{(\sigma_1^2/m) + (\sigma_2^2/n)}$

Ejemplo 9.3 (Continuación del ejemplo 9.1)

Suponga que cuando μ_1 y μ_2 (las resistencias a ceder promedio verdaderas de los dos tipos de acero) difieren cuando mucho en 5, la probabilidad de detectar tal alejamiento de H_0 (el poder de la prueba) debe ser de .90. ¿Satisface esta condición una prueba a nivel .01 con tamaños de muestra $m = 20$ y $n = 25$? El valor de σ con estos tamaños de muestra (el denominador de z) se calculó previamente como 1.34. La probabilidad de un error de tipo II con la prueba a nivel .01 de dos colas cuando $\mu_1 - \mu_2 = \Delta' = 5$ es

$$\begin{aligned}\beta(5) &= \Phi\left(2.58 - \frac{5-0}{1.34}\right) - \Phi\left(-2.58 - \frac{5-0}{1.34}\right) \\ &= \Phi(-1.15) - \Phi(-6.31) = .1251\end{aligned}$$

Es fácil verificar que también $\beta(-5) = .1251$ (porque la región de rechazo es simétrica). Por consiguiente, la probabilidad de detectar tal alejamiento es $1 - \beta(5) = .8749$. Como este valor es un poco menor que .9, se deberán utilizar tamaños de muestra un poco más grandes. ■

Como en el capítulo 8, se pueden determinar tamaños de muestra m y n que satisfagan tanto P(error de tipo I) = una α especificada y P(error de tipo II cuando $\mu_1 - \mu_2 = \Delta'$) = una β especificada. Para una prueba de cola superior, la igualación de la expresión previa para $\beta(\Delta')$ al valor especificado de β da

$$\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n} = \frac{(\Delta' - \Delta_0)^2}{(z_\alpha + z_\beta)^2}$$

Cuando los dos tamaños de muestra son iguales, esta ecuación da

$$m = n = \frac{(\sigma_1^2 + \sigma_2^2)(z_\alpha + z_\beta)^2}{(\Delta' - \Delta_0)^2}$$

Estas expresiones también son correctas para una prueba α cola inferior, en tanto α es reemplazada por $\alpha/2$ para una prueba de dos colas.

Pruebas con muestra grande

Las suposiciones de distribuciones de población normal y los valores conocidos de σ_1 y σ_2 afortunadamente son innecesarios cuando ambos tamaños de muestra son suficientemente grandes. En este caso, el teorema del límite central garantiza que $\overline{X} - \overline{Y}$ tenga de manera aproximada una distribución normal independientemente de las distribuciones de población subyacentes. Además, con S_1^2 y S_2^2 en lugar de σ_1^2 y σ_2^2 en la expresión (9.1) se obtiene una variable cuya distribución es aproximadamente normal estándar:

$$Z = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}}$$

Al reemplazar $\mu_1 - \mu_2$ con Δ_0 se obtiene un estadístico de prueba con muestra grande, el valor esperado de $\overline{X} - \overline{Y}$ cuando H_0 es verdadera. Este estadístico Z tiene aproximadamente una distribución estándar normal cuando H_0 es verdadera. Pruebas con un nivel de significancia deseado se obtienen utilizando valores críticos z exactamente como antes.

El uso del valor estadístico de prueba

$$z = \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}}$$

junto con las regiones de rechazo de colas superior, inferior y de dos colas antes formuladas basadas en valores críticos z da pruebas con muestra grande cuyos niveles de significación son aproximadamente α. Estas pruebas son apropiadas de modo normal si tanto $m > 40$ como $n > 40$. Un valor P se calcula en forma exacta como se hizo en pruebas z previas.

Ejemplo 9.4 ¿Qué impacto tiene el consumo de comida rápida en las diferentes dietas y características de salud? El artículo “Effects of Fast-Food Consumption on Energy Intake and Diet Quality Among Children in a National Household Study” (*Pediatrics*, 2004: 112–118) reportó el resumen de datos que acompaña a la ingesta diaria de calorías, tanto para una muestra de adolescentes que dijeron no suelen comer comida rápida y otra muestra de adolescentes que dijeron que acostumbran ingerir comida rápida.

Comen comida rápida	Tamaño de muestra	Media de muestra	Desv. est. muestras
No	663	2258	1519
Sí	413	2637	1138

¿Estos datos proporcionan una fuerte evidencia para concluir que la ingesta promedio de calorías real para los adolescentes que suelen ingerir comida rápida excede en más de 200 calorías por día el consumo promedio real para aquellos que normalmente no ingieren comida rápida? Vamos a investigar mediante la realización de una prueba de hipótesis en un nivel de significación de aproximadamente .05.

El parámetro de interés es $\mu_1 - \mu_2$, donde μ_1 es la ingesta media de calorías real para los adolescentes que no suelen ingerir comida rápida y μ_2 es el consumo real promedio de los adolescentes que suelen ingerir comida rápida. Las hipótesis de interés son

$$H_0: \mu_1 - \mu_2 = -200 \quad \text{versus} \quad H_a: \mu_1 - \mu_2 < -200$$

La hipótesis alternativa afirma que el verdadero consumo promedio diario de los que suelen ingerir comida rápida supera al de aquellos que no ingieren más de 200 calorías. El valor estadístico de prueba es

$$z = \frac{\bar{x} - \bar{y} - (-200)}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}}$$

La desigualdad en H_a implica que la prueba es de cola inferior; H_0 debe ser rechazada si $z \le -z_{.05} = -1.645$. El valor estadístico de prueba calculado es

$$z = \frac{2258 - 2637 + 200}{\sqrt{\frac{(1519)^2}{663} + \frac{(1138)^2}{413}}} = \frac{-179}{81.34} = -2.20$$

Como $-2.20 \le -1.645$, la hipótesis nula es rechazada. Con un nivel de significancia de .05, pareciera que el promedio real diario de consumo de calorías por adolescente quienes típicamente ingieren comida rápida excede en más de 200 el promedio real de consumo para aquellos que no ingieran comida rápida con frecuencia.

El valor P para la prueba es

$$\text{Valor } P = \text{área bajo la curva } z \text{ a la izquierda de } -2.20 = \Phi(-2.20) = .0139$$

Ya que $.0139 \le .05$, volvemos a rechazar la hipótesis nula al nivel de significación .05. Sin embargo, el valor P no es lo suficientemente pequeño como para justificar el rechazo de H_0 al nivel de significación .01.

Tenga en cuenta que si la etiqueta 1 había sido utilizada para la condición de la comida rápida y la 2 se ha utilizado para la condición de ausencia de comida rápida, entonces 200 han sustituido -200 en ambas hipótesis y H_a habría contenido la desigualdad >, lo que implica una prueba de cola superior. El valor de la prueba estadística resultante habría sido 2.20, dando el mismo valor P como antes. ■

Intervalos de confianza para $\mu_1 - \mu_2$

Cuando ambas distribuciones de población son normales la estandarización de $\bar{X} - \bar{Y}$ da una variable aleatoria Z con distribución normal estándar. Como el área bajo la curva z entre $-z_{\alpha/2}$ y $z_{\alpha/2}$ es $1 - \alpha$, se desprende que

$$P\left(-z_{\alpha/2} < \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} < z_{\alpha/2}\right) = 1 - \alpha$$

La manipulación de las desigualdades entre paréntesis para aislar $\mu_1 - \mu_2$ da la formulación de probabilidad equivalente

$$P\left(\overline{X} - \overline{Y} - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}} < \mu_1 - \mu_2 < \overline{X} - \overline{Y} + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}\right) = 1 - \alpha$$

Esto implica que un intervalo de confianza de $100(1 - \alpha)\%$ para $\mu_1 - \mu_2$ tiene un límite inferior $\bar{x} - \bar{y} - z_{\alpha/2} \cdot \sigma_{\overline{X}-\overline{Y}}$ y uno superior $\bar{x} - \bar{y} + z_{\alpha/2} \cdot \sigma_{\overline{X}-\overline{Y}}$ donde $\sigma_{\overline{X}-\overline{Y}}$ es la expresión de la raíz cuadrada. Este intervalo es un caso especial de la fórmula general $\hat{\theta} \pm z_{\alpha/2} \cdot \sigma_{\hat{\theta}}$.

Si tanto m como n son grandes, el teorema del límite central implica que este intervalo es válido incluso sin la suposición de población normal; en este caso, el intervalo de confianza es *aproximadamente* de $100(1 - \alpha)\%$. Además, el uso de las varianzas muestrales S_1^2 y S_2^2 en la variable estandarizada Z da un intervalo válido en el cual s_1^2 y s_2^2 reemplazan a σ_1^2 y σ_2^2.

Siempre que m y n son grandes, un intervalo de confianza para $\mu_1 - \mu_2$ con un nivel de confianza de aproximadamente $100(1 - \alpha)\%$ es

$$\bar{x} - \bar{y} \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}$$

donde $-$ da el límite inferior y $+$ es el límite superior del intervalo. Un límite de confianza superior o inferior también puede ser calculado reteniendo el signo apropiado ($+$ o $-$) reemplazando $z_{\alpha/2}$ con z_α.

Una regla empírica estándar para caracterizar tamaños de muestra tan grandes es $m > 40$ y $n > 40$.

Ejemplo 9.5 Un experimento realizado para estudiar varias características de pernos de anclaje arrojó 78 observaciones de resistencia al esfuerzo cortante (kip) de pernos de 3/8 pulg de diámetro y 88 observaciones de resistencia de pernos de 1/2 pulg de diámetro. A continuación se dan cantidades obtenidas con Minitab y en la figura 9.1 se presenta una gráfica de caja comparativa. Los tamaños de muestra, las medias y las desviaciones estándar muestrales concuerdan con los valores dados en el artículo "Ultimate Load Capacities of Expansion Anchor Bolts", (*J. of Energy Engr.*, 1993: 139–158). Los resúmenes sugieren que la diferencia principal entre las dos muestras es dónde están centradas.

Variable	N	Media	Mediana	MediaVerd	DesvEst	ErrorEstmedio
diam 3/8	78	4.250	4.230	4.238	1.300	0.147

Variable	Min	Max	Q1	Q3
diam 3/8	1.634	7.327	3.389	5.075

Variable	N	Media	Mediana	Media	DesvEst	ErrorEstmedio
diam 1/2	88	7.140	7.113	7.150	1.680	0.179

Variable	Min	Max	Q1	Q3
diam 1/2	2.450	11.343	5.965	8.447

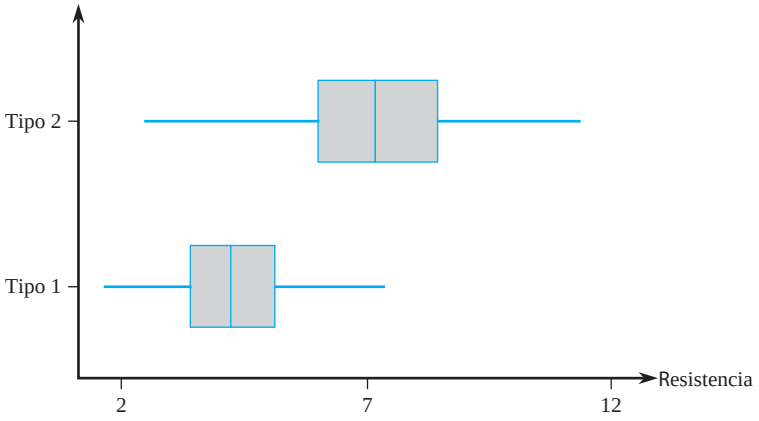


Figura 9.1 Gráfica de caja comparativa de los datos de resistencia al esfuerzo cortante

Calcule ahora un intervalo de confianza para la diferencia entre la resistencia al esfuerzo cortante promedio verdadera de pernos de 3/8 pulg (μ_1) y la resistencia al esfuerzo cortante promedio verdadera de pernos de 1/2 pulg (μ_2) con un nivel de confianza de 95%:

$$4.25 - 7.14 \pm (1.96)\sqrt{\frac{(1.30)^2}{78} + \frac{(1.68)^2}{88}} = -2.89 \pm (1.96)(.2318)$$
$$= -2.89 \pm .45 = (-3.34, -2.44)$$

Es decir, con confianza de 95%, $-3.34 < \mu_1 - \mu_2 < -2.44$. Por consiguiente se puede estar altamente confiado en que la resistencia al esfuerzo cortante verdadera de los pernos de 1/2 pulg excede la de los pernos de 3/8 pulg en entre 2.44 kips y 3.34 kips. Obsérvese que si se reetiquetan de modo que μ_1 se refiera a pernos de 1/2 pulg y μ_2 a pernos de 3/8 pulg, el intervalo de confianza ahora está centrado en $+2.89$ y el valor de .45 se sigue restando y sumando para obtener los límites de confianza. El intervalo resultante es (2.44, 3.34) y la interpretación es idéntica a la del intervalo previamente calculado. ■

Si las varianzas de σ_1^2 y σ_2^2 son por lo menos aproximadamente conocidas y el investigador utiliza tamaños de muestra iguales, entonces el tamaño de muestra común n que da un intervalo de $100(1 - \alpha)\%$ de ancho w es

$$n = \frac{4z_{\alpha/2}^2(\sigma_1^2 + \sigma_2^2)}{w^2}$$

la que normalmente tiene que ser redondeada a un entero.

EJERCICIOS Sección 9.1 (1–16)

1. Un artículo que apareció en el ejemplar de noviembre de 1983 del *Consumer Reports* comparó varios tipos de baterías. Las duraciones promedio de baterías AA alcalinas Duracell y alcalinas Eveready Energizer se dieron como 4.1 horas y 4.5 horas, respectivamente. Suponga que éstas son las duraciones promedio de la población.

a. Sea $\overline{X}$ la duración promedio muestral de 100 baterías Duracell y $\overline{Y}$ la duración promedio muestral de 100 baterías Eveready. ¿Cuál es el valor medio $\overline{X} - \overline{Y}$ (es decir, dónde está centrada la distribución de $\overline{X} - \overline{Y}$)? ¿Cómo depende su respuesta de los tamaños de muestra especificados?

b. Suponga que la desviación estándar de población de duración es de 1.8 horas para baterías Duracell y de 2.0 horas para baterías Eveready. Con los tamaños de muestra dados en el inciso (a), ¿cuál es la varianza del estadístico $\overline{X} - \overline{Y}$ y cuál es su desviación estándar?

c. Con los tamaños de muestra dados en el inciso (a) trace la curva de distribución aproximada de $\overline{X} - \overline{Y}$ (incluya una escala de medición sobre el eje horizontal). ¿Sería la forma de la curva necesariamente la misma con tamaños de muestra de 10 baterías de cada tipo? Explique.

2. La National Health Statistics Reports en los informes de fecha 22 de octubre de 2008, incluye la siguiente información sobre la altura (pulgadas) para las mujeres blancas no hispanas.

Edad	Tamaño muestral	Media muestral	Error estándar de la media
20–39	866	64.9	.09
60 y más	934	63.1	.11

a. Calcule e interprete un intervalo de confianza al nivel de confianza de aproximadamente el 95% de la diferencia entre la altura media de la población de las mujeres más jóvenes y las mujeres mayores.

b. Sea μ_1 que denota la media poblacional de altura para las personas de 20 a 39 años y μ_2 denota la media poblacional de altura para las mayores de 60 años. Interprete las hipótesis H_0: $\mu_1 - \mu_2 = 1$ y H_a: $\mu_1 - \mu_2 > 1$ y, a continuación, lleve a cabo una prueba de estas hipótesis al nivel de significación .001 utilizando el enfoque de región de rechazo.

c. ¿Cuál es el valor P para la prueba que llevó a cabo en el inciso (b)? Con base en este valor P, ¿se rechaza la hipótesis nula a cualquier nivel de significancia razonable? Explique.

d. ¿Qué hipótesis sería conveniente si μ_1 se refiere al grupo de mayor edad, μ_2 al grupo de edad más joven y desea ver si había pruebas de peso para concluir que la población de la altura media de las mujeres más jóvenes superó al de las mujeres de mayor edad en más de 1 pulgada?

3. Sea μ_1 la duración de la banda de rodamiento promedio verdadera de una marca premium de neumático radial P205/65R15 y sea μ_2 la duración de la banda de rodamiento promedio verdadera de una marca económica de un neumático de la misma medida. Pruebe H_0: $\mu_1 - \mu_2 = 5000$ contra H_a: $\mu_1 - \mu_2 > 5000$ a un nivel .01 con los siguientes datos: $m = 45$, $\bar{x} = 42{,}500$, $s_1 = 2200$, $n = 45$, $\bar{y} = 36{,}800$ y $s_2 = 1500$.

4. a. Use los datos del ejemplo 9.4 para calcular un intervalo de confianza de 95% para $\mu_1 - \mu_2$. ¿Sugiere el intervalo resultante que $\mu_1 - \mu_2$ ha sido estimado con precisión?

b. Use los datos del ejercicio 3 para calcular un límite de confianza superior de 95% para $\mu_1 - \mu_2$.

5. Las personas que padecen el síndrome de Reynaud están propensas a sufrir un deterioro repentino de la circulación sanguínea en los dedos de las manos y de los pies. En un experimento para estudiar el grado de este deterioro, cada uno de los sujetos sumergió un dedo índice en agua y se midió la producción de calor resultante (cal/cm^2/min). Con $m = 10$ sujetos con el síndrome, la producción de calor promedio fue $\bar{x} = .64$ y con $n = 10$ sin el síndrome, la producción promedio fue de 2.05. Sean μ_1 y μ_2 las producciones de calor promedio verdaderas de los dos tipos de sujetos. Suponga que las dos distribuciones de producción de calor son normales con $\sigma_1 = .2$ y $\sigma_2 = .4$.

a. Considere probar H_0: $\mu_1 - \mu_2 = -1.0$ contra H_a: $\mu_1 - \mu_2 < -1.0$ a un nivel .01. Describa en palabras qué dice H_a y luego realice la prueba.

b. Calcule el valor P para el valor de Z obtenido en el inciso (a).

c. ¿Cuál es la probabilidad de un error de tipo II cuando la diferencia real entre μ_1 y μ_2 es $\mu_1 - \mu_2 = -1.2$?

d. Suponiendo que $m = n$, ¿qué tamaños de muestra se requieren para asegurar que $\beta = .1$ cuando $\mu_1 - \mu_2 = -1.2$?

6. Un experimento para comparar la resistencia de adhesión a la tensión de un mortero modificado con un polímero de látex (mortero de cemento Portland al cual se le agregan emulsiones de polímero de látex durante la mezcla) con la de un mortero no modificado dio por resultado $\bar{x} = 18.12$ kgf/cm^2 para el mortero modificado ($m = 40$) y $\bar{y} = 16.87$ kgf/cm^2 para el mortero no modificado ($n = 32$). Sean μ_1 y μ_2 las resistencias de adhesión a la tensión promedio verdaderas para los morteros modificado y no modificado, respectivamente. Suponga que ambas distribuciones de la resistencia de adhesión son normales.

a. Suponiendo que $\sigma_1 = 1.6$ y $\sigma_2 = 1.4$, pruebe H_0: $\mu_1 - \mu_2 = 0$ contra H_a: $\mu_1 - \mu_2 > 0$ a un nivel .01.

b. Calcule la probabilidad de un error de tipo II para la prueba del inciso (a) cuando $\mu_1 - \mu_2 = 1$.

c. Suponga que el investigador decidió utilizar una prueba a un nivel .05 y deseaba $\beta = .10$ cuando $\mu_1 - \mu_2 = 1$. Si $m = 40$, ¿qué valor de n es necesario?

d. ¿Cómo cambiaría el análisis y conclusión del inciso (a) si σ_1 y σ_2 fueran desconocidas pero $s_1 = 1.6$ y $s_2 = 1.4$?

7. ¿Hay alguna tendencia sistemática de los profesores universitarios de tiempo parcial para sujetar a sus estudiantes a las diferentes normas que hacen los profesores de tiempo completo? El artículo "Are There Instructional Differences Between Full-Time and Part-Time Faculty?" (*College Teaching*, 2009: 23–26) informó que para una muestra de 125 cursos impartidos por profesores de tiempo completo, la media de GPA fue 2.7186 y la desviación estándar fue de .63342, mientras que para una muestra de 88 cursos impartidos por trabajadores de tiempo parcial, la media y desviación estándar fueron 2.8639 y .49241, respectivamente. ¿Parece que el promedio verdadero del curso de GPA para el profesorado de tiempo parcial difiere del de tiempo completo? Pruebe la hipótesis apropiada al nivel de significación .01, para obtener primero un valor P.

8. Se realizaron pruebas de resistencia a la tensión en dos grados diferentes de alambrón ("Fluidized Bed Patenting of Wire Rods", *Wire J.*, junio de 1977: 56–61) y se obtuvieron los datos adjuntos.

Grado	Tamaño muestral	Media muestral (kg/mm^2)	Desv. est. muestral
AISI 1064	$m = 129$	$\bar{x} = 107.6$	$s_1 = 1.3$
AISI 1078	$n = 129$	$\bar{y} = 123.6$	$s_2 = 2.0$

a. ¿Proporcionan los datos evidencia precisa para concluir que la resistencia promedio verdadera del grado 1078 excede la del grado 1064 en más de 10 kg/mm^2? Pruebe las hipótesis apropiadas con el método del valor P.

b. Estime la diferencia entre resistencias promedio verdaderas para los dos grados en una forma que proporcione información sobre precisión y confiabilidad.

9. El artículo "Evaluation of a Ventilation Strategy to Prevent Barotrauma in Patients at High Risk for Acute Respiratory Distress Syndrome" (*New Engl. J. of Med.*, 1998: 355–358) reportó sobre un experimento en el cual 120 pacientes con características clínicas similares fueron divididos al azar en un grupo de control y un grupo de tratamiento, cada uno compuesto de 60 pacientes. La permanencia en la unidad de cuidados intensivos

media UCI y la desviación estándar muestrales para el grupo de tratamiento fueron 19.9 y 39.1, respectivamente, en tanto que estos valores para el grupo de control fueron 13.7 y 15.8.

a. Calcule una estimación puntual de la diferencia entre la permanencia en la unidad de cuidados intensivos promedio verdadera para los grupos de tratamiento y control. ¿Sugiere esta estimación que existe una diferencia significativa entre las permanencias promedio verdaderas en las dos condiciones?

b. Responda la pregunta planteada en el inciso (a) realizando una prueba formal de hipótesis. ¿Es diferente el resultado de lo que había conjeturado en el inciso (a)?

c. ¿Parece que la permanencia en la unidad de cuidados intensivos de pacientes a los que se les administró tratamiento de ventilación está normalmente distribuida? Explique su razonamiento.

d. Estime el tiempo de permanencia promedio verdadero de pacientes a los que se les administró tratamiento de ventilación en una forma que dé información sobre precisión y confiabilidad.

10. Se realizó un experimento para comparar la tenacidad a la fractura de acero maraging de alta pureza con 18% de níquel con acero de pureza comercial del mismo tipo (*Corrosion Science*, 1971: 723–736). Con $m = 32$ especímenes, la tenacidad promedio muestral fue $\bar{x} = 65.6$ para el acero de alta pureza, en tanto que para $n = 38$ especímenes de acero comercial $\bar{y} = 59.8$. Como el acero de alta pureza es más caro, su uso en ciertas aplicaciones se justifica sólo si su tenacidad a la fractura supera la del acero de pureza comercial en más de 5. Suponga que ambas distribuciones de tenacidad son normales.

a. Suponiendo que $\sigma_1 = 1.2$ y $\sigma_2 = 1.1$, pruebe las hipótesis pertinentes con $\alpha = .001$.

b. Calcule β para la prueba realizada en el inciso (a) cuando $\mu_1 - \mu_2 = 6$.

11. Se determinó el nivel de plomo en la sangre con una muestra de 152 trabajadores de desechos peligrosos de 20 a 30 años de edad y también con una muestra de 86 trabajadoras y el resultado fue un error estándar medio $\pm$ de 5.5 $\pm$ 0.3 para los hombres y de 3.8 $\pm$ 0.2 para las mujeres (“Temporal Changes in Blood Lead Levels of Hazardous Waste Workers in New Jersey, 1984–1987”, *Environ. Monitoring and Assessment*, 1993: 99–107). Estime la diferencia entre niveles de plomo en sangre promedio verdaderos para trabajadores y trabajadoras en una forma que proporcione información sobre confiabilidad y precisión.

12. La tabla adjunta contiene datos sobre resistencia a la compresión (N/mm^2) de especímenes de concreto hechos con una mezcla de cenizas combustibles pulverizadas (“A Study of Twenty-Five-Year-Old Pulverized Fuel Ash Concrete Used in Foundation Structures”, *Proc. Inst. Civ. Engrs.*, marzo de 1985: 149–165):

Edad (días)	Tamaño muestral	Media muestral	Desv. est. muestral
7	68	26.99	4.89
28	74	35.76	6.43

Calcule e interprete un intervalo de confianza de 99% para la diferencia entre resistencia a 7 días promedio verdadera y resistencia a 28 días promedio verdadera.

13. Un ingeniero mecánico desea comparar las propiedades de resistencia de vigas de acero con vigas similares hechas de una aleación particular. Se probará el mismo número de vigas, *n*, de cada tipo. Cada viga se colocará en posición horizontal con un apoyo en cada extremo y se aplicará una fuerza de 2500 lb en el centro y se medirá la deflexión. Por experiencias pasadas con las mismas vigas, el ingeniero desea suponer que la desviación estándar verdadera de la deflexión de ambos tipos de viga es de .05 pulg. Como la aleación es más cara, el ingeniero desea probar a un nivel de .01 si su deflexión promedio es más pequeña que la de la viga de acero. ¿Qué valor de *n* es apropiado si la probabilidad de error de tipo II deseado es de .05 cuando la diferencia de deflexión promedio verdadera favorece la aleación por .04 pulg?

14. Se determinó el nivel de actividad de oxidasa monoamina (MAO, por su siglas en inglés) en plaquetas sanguíneas (nm/mg proteína/h) para cada individuo en una muestra de 43 esquizofrénicos crónicos y el resultado fue $\bar{x} = 2.69$ y $s_1 = 2.30$, así como también para 45 sujetos normales y el resultado fue $\bar{y} = 6.35$ y $s_2 = 4.03$. ¿Sugieren fuertemente estos datos que la actividad de MAO promedio verdadera en sujetos normales es más de dos veces que el nivel de actividad en esquizofrénicos? Obtenga un procedimiento de prueba y realice una prueba con $\alpha = .01$. [*Sugerencia*: H_0 y H_a en este caso tienen una forma diferente de los tres casos estándar. Si μ_1 y μ_2 se refieren a la actividad de MAO promedio verdadera para sujetos esquizofrénicos y normales, respectivamente, considere el parámetro $\theta = 2\mu_1 - \mu_2$. Escriba H_0 y H_a en función de θ, estime θ y derive $\hat{\sigma}_{\hat{\theta}}$ (“Reduced Monoamine Oxidase Activity in Blood Platelets from Schizophrenic Patients”, *Nature*, 28 de julio de 1972: 225–226).]

15. **a.** Demuestre que para la prueba de cola superior con σ_1 y σ_2 conocidas a medida que m o n se incrementa, β disminuye cuando $\mu_1 - \mu_2 > \Delta_0$.

b. En el caso de tamaños de muestra iguales ($m = n$) y α fijo, ¿qué le sucede al tamaño de muestra necesario n a medida que β disminuye, donde β es la probabilidad de error de tipo II deseada con una alternativa fija?

16. Para decidir si dos tipos diferentes de acero tienen los mismos valores de tenacidad a la fractura promedio verdaderos, se probaron n especímenes de cada tipo y se obtuvieron los siguientes resultados:

Tipo	Promedio muestral	Desv. est. muestral
1	60.1	1.0
2	59.9	1.0

Calcule el valor P para la prueba z con dos muestras apropiadas, suponiendo que los datos se basaron en $n = 100$. Luego repita el cálculo con $n = 400$. ¿Es el valor P pequeño con $n = 400$ indicativo de una diferencia que tenga significación práctica? ¿Se sentiría satisfecho con sólo el reporte del valor P? Comente brevemente.

9.2 Prueba *t* con dos muestras e intervalo de confianza

Normalmente un investigador no conoce los valores de las varianzas de población. En la sección previa, se ilustró por lo que se refiere a muestras grandes el uso de un procedimiento de prueba *z* y un intervalo de confianza en el cual se utilizaron las varianzas muestrales en lugar de las varianzas de población. En realidad, con muestras grandes, el teorema del límite central permite utilizar estos métodos incluso cuando las dos poblaciones de interés no son normales.

En la práctica, sin embargo, a menudo sucede que por lo menos un tamaño de muestra es pequeño y las variaciones de población tienen valores desconocidos. Sin el TLC a nuestra disposición, se procede haciendo suposiciones específicas sobre las distribuciones de población subyacentes. El uso de procedimientos inferenciales que se desprenden de estas suposiciones se limita entonces a situaciones en las que las suposiciones se satisfacen por lo menos de forma aproximada. Podríamos, por ejemplo, suponer que las dos distribuciones de la población son miembros de la familia de Weibull o que ambas son distribuciones de Poisson. No debería sorprenderle el enterarse de que la normalidad es generalmente la hipótesis más razonable.

SUPOSICIONES

Ambas distribuciones de población son normales, de modo que $X_1, X_2, \ldots, X_m$ es una muestra aleatoria de una distribución normal y por tanto es $Y_1, \ldots, Y_n$ (con las X y Y independientes entre sí). La factibilidad de estas suposiciones puede ser juzgada construyendo una gráfica de probabilidad normal de las x_i y otra de las y_i.

El estadístico de prueba y la fórmula del intervalo de confianza están basados en la misma variable estandarizada desarrollada en la sección 9.1, pero la distribución pertinente ahora es *t* en lugar de *z*.

TEOREMA

Cuando ambas distribuciones de población son normales, la variable estandarizada

$$T = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}} \tag{9.2}$$

tiene aproximadamente una distribución *t* con ν grados de libertad estimados a partir de los datos como sigue

$$\nu = \frac{\left(\frac{s_1^2}{m} + \frac{s_2^2}{n}\right)^2}{\frac{(s_1^2/m)^2}{m-1} + \frac{(s_2^2/n)^2}{n-1}} = \frac{[(se_1)^2 + (se_2)^2]^2}{\frac{(se_1)^4}{m-1} + \frac{(se_2)^4}{n-1}}$$

donde

$$se_1 = \frac{s_1}{\sqrt{m}}, se_2 = \frac{s_2}{\sqrt{n}}$$

(redondear ν al entero más cercano hacia abajo).

La manipulación de T en un enunciado de probabilidad para aislar $\mu_1 - \mu_2$ da un intervalo de confianza, en tanto que al reemplazar $\mu_1 - \mu_2$ con el valor nulo Δ_0 se obtiene un estadístico de prueba.

El **intervalo de confianza *t* con dos muestras para $\mu_1 - \mu_2$** con nivel de confianza de $100(1-\alpha)\%$ es entonces

$$\bar{x} - \bar{y} \pm t_{\alpha/2,\nu}\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}$$

Se puede calcular un límite de confianza unilateral como se describió con anterioridad.

La **prueba *t* con dos muestras** para probar H_0: $\mu_1 - \mu_2 = \Delta_0$ es como sigue:

$$\text{Valor estadístico de prueba: } t = \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}}$$

Hipótesis alternativa	Región de rechazo con una prueba a nivel α aproximado
H_a: $\mu_1 - \mu_2 > \Delta_0$	$t \geq t_{\alpha,\nu}$ (cola superior)
H_a: $\mu_1 - \mu_2 < \Delta_0$	$t \leq -t_{\alpha,\nu}$ (cola inferior)
H_a: $\mu_1 - \mu_2 \neq \Delta_0$	$t \geq t_{\alpha/2,\nu}$ o $t \leq -t_{\alpha/2,\nu}$ (dos colas)

Se puede calcular un valor *P* como se describió en la sección 8.4 para la prueba *t* con una muestra.

Ejemplo 9.6 El volumen de huecos en una tela afecta las propiedades de comodidad, flamabilidad y aislantes. La permeabilidad de una tela se refiere a la accesibilidad de los espacios huecos al flujo de un gas o líquido. El artículo "The Relationship Between Porosity and Air Permeability of Woven Textile Fabrics" (*J. of Testing and Eval.*, 1997: 108-114) contiene información resumida sobre permeabilidad al aire ($cm^3/cm^2/s$) de varios tipos diferentes de tela. Considere los siguientes datos sobre dos tipos diferentes de tela de tejido ordinario:

Tipo de tela	Tamaño de muestra	Media de la muestra	Desviación estándar de la muestra
Algodón	10	51.71	.79
Triacetato	10	136.14	3.59

Suponiendo que las distribuciones de porosidad de ambos tipos de tela son normales, calcule un intervalo de confianza para la diferencia entre la porosidad promedio verdadera de la tela de algodón y la de la tela de acetato, utilizando un nivel de confianza de 95%. Antes de que se pueda seleccionar un valor crítico *t* apropiado, se debe determinar el número de grados de libertad:

$$\text{gl} = \frac{\left(\frac{.6241}{10} + \frac{12.8881}{10}\right)^2}{\frac{(.6241/10)^2}{9} + \frac{(12.8881/10)^2}{9}} = \frac{1.8258}{.1850} = 9.87$$

Así pues se utiliza $\nu = 9$; la tabla A.5 del apéndice da $t_{.025,9} = 2.262$. El intervalo resultante es

$$51.71 - 136.14 \pm (2.262)\sqrt{\frac{.6241}{10} + \frac{12.8881}{10}} = -84.43 \pm 2.63$$
$$= (-87.06, -81.80)$$

Con un alto grado de confianza, se puede decir que la porosidad promedio verdadera de especímenes de tela de triacetato excede la de los especímenes de algodón por entre 81.80 y 87.06 $cm^3/cm^2/s$. ■

Ejemplo 9.7 El deterioro de muchas redes de tuberías municipales a través del país es una preocupación creciente. Una tecnología propuesta para rehabilitar las tuberías utiliza un forro flexible insertado en las tuberías existentes. El artículo "Effect of Welding on a High-Density Polyethylene Liner" (*J. of Materials in Civil Engr.*, 1996: 94–100) reportó los siguientes datos de resistencia a la tensión ($lb/pulg^2$) de especímenes de forro cuando se utilizó cierto proceso de fusión y cuando este proceso no se utilizó.

Sin fusión	2748	2700	2655	2822	2511			
	3149	3257	3213	3220	2753			
	$m = 10$	$\bar{x} = 2902.8$		$s_1 = 277.3$				
Fusionado	3027	3356	3359	3297	3125	2910	2889	2902
	$n = 8$	$\bar{y} = 3108.1$		$s_2 = 205.9$				

La figura 9.2 muestra curvas de probabilidad normal generadas por Minitab. El patrón lineal de cada una confirma la suposición de que las dos distribuciones de resistencia a la tensión en las dos condiciones son normales.

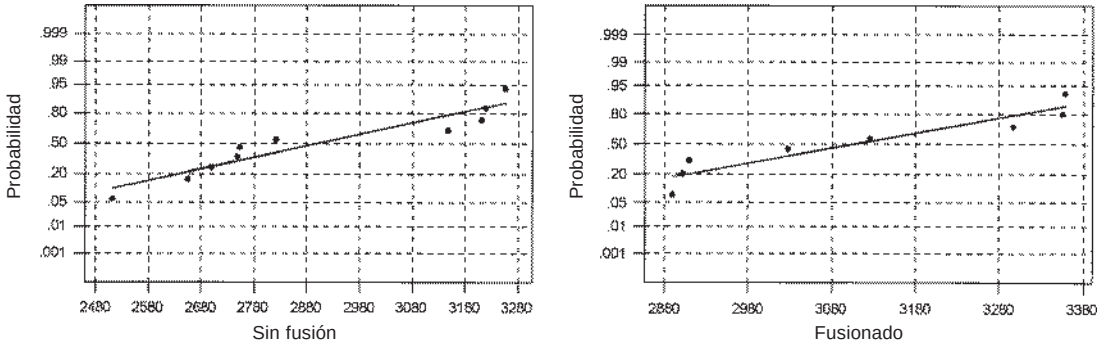


Figura 9.2 Gráficas de probabilidad normal con Minitab para los datos de resistencia a la tensión

Los autores del artículo afirman que el proceso de fusión incrementó la resistencia a la tensión promedio. El mensaje de la gráfica de caja comparativa de la figura 9.3 no es del todo claro. Realice una prueba de hipótesis para ver si los datos confirman esta conclusión.

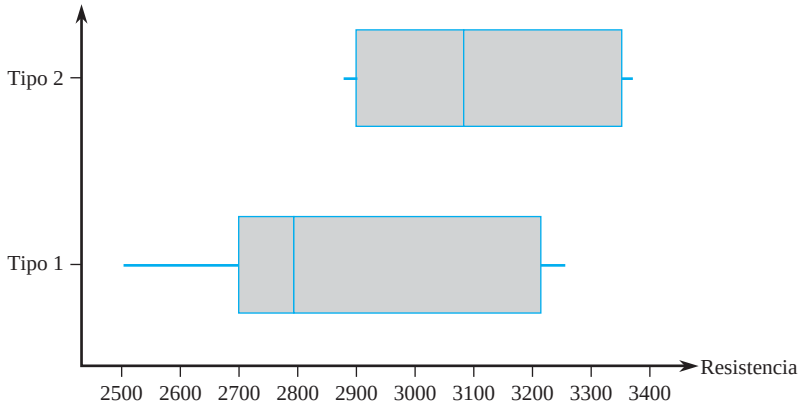


Figura 9.3 Gráfica de caja comparativa de los datos de resistencia a la tensión

1. Sea μ_1 la resistencia a la tensión promedio verdadera de especímenes cuando se utiliza el tratamiento de no fusión y μ_2 la resistencia a la tensión promedio verdadera cuando se utiliza el tratamiento de fusión.
2. H_0: $\mu_1 - \mu_2 = 0$ (ninguna diferencia en las resistencias a la tensión promedio verdaderas con los dos tratamientos)
3. H_a: $\mu_1 - \mu_2 < 0$ (la resistencia a la tensión promedio verdadera del tratamiento sin fusión es menor que la del tratamiento de fusión, de modo que la conclusión de los investigadores es correcta)
4. El valor nulo es $\Delta_0 = 0$, de modo que el valor del estadístico de prueba es

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}}$$

5. A continuación se calcula tanto el valor estadístico de prueba como el número de grados de libertad para la prueba:

$$t = \frac{2902.8 - 3108.1}{\sqrt{\frac{(277.3)^2}{10} + \frac{(205.9)^2}{8}}} = \frac{-205.3}{113.97} = -1.8$$

Con $s_1^2/m = 7689.529$ y $s_2^2/n = 5299.351$,

$$\nu = \frac{(7689.529 + 5299.351)^2}{(7689.529)^2/9 + (5299.351)^2/7} = \frac{168{,}711{,}003.7}{10{,}581{,}747.35} = 15.94$$

así, la prueba se basará en 15 grados de libertad.

6. La tabla A.8 del apéndice muestra que el área bajo la curva t con 15 grados de libertad a la derecha de 1.8 es .046, de modo que el valor P para una prueba de cola inferior también es .046. Los siguientes datos generados por Minitab resumen todos los cálculos:

```
T muestreados para no fusión vs fusionada

             N    Media   DesvEs     MediaSE
sin fusión  10     2903      277          88
fusionado    8     3108      206          73

95% IC para mu no fusionado-mu fusionado: (-488, 38)
Prueba T mu no fusionada = mu fusionada (vs <): T = -1.80 P = 0.046 DF = 15
```

7. Con un nivel de significancia de .05, apenas si se puede rechazar la hipótesis nula a favor de la hipótesis alternativa, lo que confirma la conclusión expresada en el artículo. No obstante, alguien que demande evidencia más contundente podría seleccionar $\alpha = .01$, un nivel con el cual H_0 no puede ser rechazada.

Si la pregunta planteada hubiera sido si la fusión incrementó la resistencia promedio verdadera en más de 100 lb/pulg^2, entonces las hipótesis pertinentes habrían sido H_0: $\mu_1 - \mu_2 = -100$ contra H_a: $\mu_1 - \mu_2 < -100$; es decir, el valor nulo habría sido $\Delta_0 = -100$. ■

Procedimientos *t* agrupados

De la suposición de que no sólo las dos distribuciones de población son normales sino que también tienen varianzas iguales ($\sigma_1^2 = \sigma_2^2$) se deriva la alternativa de los procedimientos t con dos muestras. Es decir, las dos curvas de distribución de población se suponen normales con dispersiones iguales, la única diferencia posible entre ellas sería dónde están centradas.

Sea σ^2 la varianza de población común. Luego, estandarizando $\overline{X} - \overline{Y}$ se obtiene

$$Z = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2}{m} + \frac{\sigma^2}{n}}} = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\sigma^2\left(\frac{1}{m} + \frac{1}{n}\right)}}$$

cuya distribución es normal estándar. Antes de que esta variable pueda ser utilizada como base para hacer inferencias con respecto a $\mu_1 - \mu_2$, se debe estimar la varianza común a partir de los datos muestrales. Un estimador de σ^2 es S_1^2, la varianza de las m observaciones en la primera muestra y otro es S_2^2, la varianza de la segunda muestra. Intuitivamente, se obtiene un mejor estimador que cualquier varianza muestral individual al combinar las dos varianzas muestrales. Un primer intento podría ser utilizar $(S_1^2 + S_2^2)/2$. No obstante, si $m > n$, entonces la primera muestra contiene más información sobre σ^2 que la segunda y un comentario análogo es válido si $m < n$. El siguiente promedio *ponderado* de las dos varianzas muestrales, llamado **estimador agrupado** (es decir, combinado) **de σ^2,** se ajusta a cualquier diferencia que exista entre los dos tamaños de muestra:

$$S_p^2 = \frac{m - 1}{m + n - 2} \cdot S_1^2 + \frac{n - 1}{m + n - 2} \cdot S_2^2$$

La primera muestra contribuye con $m - 1$ grados de libertad a la estimación de σ^2 y la segunda con $n - 1$ grados de libertad, para un total de $m + n - 2$ grados de libertad. La teoría estadística dice que si S_p^2 reemplaza a σ^2 en la expresión para Z, la variable estandarizada resultante tiene una distribución t basada en $m + n - 2$ grados de libertad. Del mismo modo que las variables estandarizadas con anterioridad se utilizaron como base para deducir intervalos de confianza y procedimientos de prueba, esta variable t conduce de inmediato al intervalo de confianza t agrupado para estimar $\mu_1 - \mu_2$ y a la prueba t agrupada para probar hipótesis con respecto a una diferencia entre las medias.

En el pasado, muchos estadísticos recomendaban estos procedimientos t agrupados sobre los procedimientos t con dos muestras. La prueba t agrupada, por ejemplo, puede deducirse del principio de razón de probabilidad, mientras que la prueba t con dos muestras no es una prueba de razón de probabilidad. Además, el nivel de significación para la prueba t agrupada es exacto, en tanto que sólo es aproximado para la prueba t con dos muestras. Sin embargo, investigaciones recientes han demostrado que aunque la prueba t agrupada supera por poco el desempeño de la prueba t con dos muestras (las β más pequeñas con la misma α) cuando $\sigma_1^2 = \sigma_2^2$, la primera prueba puede llevar fácilmente a conclusiones erróneas si se aplica cuando las varianzas son diferentes. Comentarios análogos se aplican al comportamiento de los dos intervalos de confianza. Es decir, los procedimientos t agrupados no son robustos a violaciones de la suposición de varianza igual.

Se ha sugerido que se podría realizar una prueba preliminar de H_0: $\sigma_1^2 = \sigma_2^2$ y utilizar un procedimiento t agrupado si esta hipótesis nula no es rechazada. Desafortunadamente, la "prueba F" usual de varianzas iguales (sección 9.5) es bastante sensible a la suposición de distribuciones de población normales, mucho más que los procedimientos t. Por consiguiente, se recomienda el método conservador de utilizar procedimientos t con dos muestras a menos que exista evidencia realmente contundente para proceder de otra manera, en particular cuando los dos tamaños de muestra son diferentes.

Probabilidades de error de tipo II

La determinación de probabilidades de error de tipo II (o de forma equivalente, potencia $= 1 - \beta$) con la prueba t con dos muestras es complicada. Parece que no existe una forma simple de utilizar las curvas β de la tabla A.17 del apéndice. La versión más reciente de Minitab (versión 16) calculará potencia para la prueba t agrupada pero no para la prueba t con dos muestras.

Sin embargo, la página de inicio del Departamento de Estadística de la UCLA (http://www.stat.ucla.edu) permite el acceso a una calculadora de potencia que llevará a cabo esto. Por ejemplo, se especificó $m = 10$, $n = 8$, $\sigma_1 = 300$ y $\sigma_2 = 225$ (éstos son los tamaños de muestra del ejemplo 9.7, cuyas desviaciones estándar muestrales son algo más pequeñas que estos valores de σ_1 y σ_2) y demandan la potencia de una prueba con nivel de .05 de dos colas de H_0: $\mu_1 - \mu_2 = 0$ cuando $\mu_1 - \mu_2 = 100$, 250 y 500. Los valores de la potencia obtenidos son .1089, .4609 y .9635 (correspondientes a $\beta = .89$, .54 y .04), respectivamente). En general, β disminuirá a medida que se incrementan los tamaños de muestra, a medida que se incrementa α y a medida que $\mu_1 - \mu_2$ se aleja de 0. El programa también calculará los tamaños de muestra necesarios a fin de obtener un valor específico de potencia para un valor particular de $\mu_1 - \mu_2$.

EJERCICIOS Sección 9.2 (17–35)

17. Determine el número de grados de libertad para la prueba t con dos muestras o el intervalo de confianza en cada una de las siguientes situaciones:
a. $m = 10, n = 10, s_1 = 5.0, s_2 = 6.0$
b. $m = 10, n = 15, s_1 = 5.0, s_2 = 6.0$
c. $m = 10, n = 15, s_1 = 2.0, s_2 = 6.0$
d. $m = 12, n = 24, s_1 = 5.0, s_2 = 6.0$

18. Sean μ_1 y μ_2 las densidades promedio verdaderas de dos tipos diferentes de ladrillos. Suponiendo normalidad de las dos distribuciones de densidad, pruebe H_0: $\mu_1 - \mu_2 = 0$ contra H_a: $\mu_1 - \mu_2 \neq 0$ con los siguientes datos: $m = 6$, $\bar{x} = 22.73$, $s_1 = .164$, $n = 5$, $\bar{y} = 21.95$ y $s_2 = .240$.

19. Suponga que μ_1 y μ_2 son distancias de frenado medias verdaderas a 50 mph de automóviles de cierto tipo equipados con dos tipos diferentes de sistemas de frenos. Use la prueba t con dos muestras a un nivel de significación de .01 para demostrar H_0: $\mu_1 - \mu_2 = -10$ contra H_a: $\mu_1 - \mu_2 < -10$ con los siguientes datos: $m = 6, \bar{x} = 115.7, s_1 = 5.03, n = 6, \bar{y} = 129.3$, y $s_2 = 5.38$.

20. Use los datos del ejercicio 19 para calcular un intervalo de confianza de 95% para la diferencia entre distancia de frenado promedio verdadera de automóviles equipados con el sistema 1 y automóviles equipados con el sistema 2. ¿Sugiere el intervalo que está disponible información precisa sobre el valor de esta diferencia?

21. Se requieren técnicas no invasivas cuantitativas para la valoración rutinaria de neuropatías periféricas, tales como el síndrome de túnel carpiano (CTS, por sus siglas en inglés). El artículo "A Gap Detection Tactility Test for Sensory Deficits Associated with Carpal Tunnel Syndrome" (*Ergonomics*, 1995: 2588–2601) reportó sobre una prueba que implicaba detectar una pequeña grieta en una superficie en otras circunstancias lisa tentando con un dedo; esto funcionalmente se asemeja a muchas actividades táctiles relacionadas con el trabajo, tal como detectar rasguños o defectos superficiales. Cuando no se permitía tentar con los dedos, el umbral de detección de grietas promedio muestral con $m = 8$ sujetos normales fue de 1.71 mm y la desviación estándar de la media .53 y con $n = 10$ sujetos con el síndrome de túnel carpiano, la media y desviación estándar muestrales fueron 2.53 y .87, respectivamente. ¿Sugieren estos datos que el umbral de detección de grietas promedio verdadero de sujetos con CTS excede el de sujetos normales? Formule y pruebe las hipótesis pertinentes utilizando un nivel de significación de .01.

22. La prueba de esfuerzo cortante sesgado es ampliamente aceptada para evaluar la adhesión de materiales de reparación resinosos para concreto; utiliza especímenes cilíndricos de dos mitades idénticas adheridas a 30°. El artículo "Testing the Bond Between Repair Materials and Concrete Substrate" (*ACI Materials J.*, 1996: 553–558) reportó que para 12 especímenes preparados utilizando un cepillo de alambre, la resistencia al esfuerzo cortante media (N/mm^2) y la desviación estándar muestral fueron de 19.20 y 1.58, respectivamente, mientras que para 12 especímenes cincelados a mano fueron de 23.13 y 4.01. ¿Parece ser diferente la resistencia promedio verdadera con los dos métodos diferentes de preparación de la superficie? Formule y pruebe las hipótesis pertinentes con un nivel de significación de .05. ¿Qué está suponiendo sobre las distribuciones del esfuerzo cortante?

23. Se están utilizando forros internos fusibles con creciente frecuencia para soportar la tela externa y mejorar la forma y caída de varias piezas de ropa. El artículo "Compatibility of Outer and Fusible Interlining Fabrics in Tailored Garments" (*Textile Res. J.*, 1997: 137–142) dio los datos adjuntos sobre extensibilidad (%) a 100 g/cm tanto de especímenes de telas de alta calidad (H) como especímenes de telas de baja calidad (P).

H	1.2	.9	.7	1.0	1.7	1.7	1.1	.9	1.7
	1.9	1.3	2.1	1.6	1.8	1.4	1.3	1.9	1.6
	.8	2.0	1.7	1.6	2.3	2.0			
P	1.6	1.5	1.1	2.1	1.5	1.3	1.0	2.6	

a. Construya gráficas de probabilidad normal para verificar la factibilidad con ambas muestras seleccionadas de distribuciones de población normales.
b. Construya una gráfica de caja comparativa. ¿Sugiere ésta que existe una diferencia entre la extensibilidad promedio verdadera de especímenes de tela de alta calidad y la de especímenes de baja calidad?
c. La media y desviación estándar muestrales de la muestra de alta calidad son 1.508 y .444, respectivamente, y las de la

muestra de baja calidad son 1.588 y .530. Use la prueba *t* con dos muestras para decidir si la extensibilidad promedio verdadera difiere para los dos tipos de tela.

24. Los daños en uvas a causa de la depredación de pájaros es un problema serio para los viticultores. El artículo "Experimental Method to Investigate and Monitor Bird Behavior and Damage to Vineyards" (*Amer. J. of Enology and Viticulture*, 2004: 288–291) reportó sobre un experimento que implica una mesa alimentadora de pájaros, un video del tiempo transcurrido y alimentos artificiales. Se recopiló información para dos especies de pájaros diferentes tanto en el sitio experimental como en un entorno de viñedo natural. Considere los siguientes datos de tiempo (s) empleado en una sola visita al lugar.

Especie	Ubicación	n	$\bar{x}$	DE media
Mirlos	Experimental	65	13.4	2.05
Mirlos	Natural	50	9.7	1.76
Silverreyes	Experimental	34	49.4	4.78
Silverreyes	Natural	46	38.4	5.06

a. Calcule un límite de confianza superior para el tiempo promedio verdadero que los mirlos emplean en una sola visita en el lugar experimental.

b. ¿Parece que el tiempo promedio verdadero empleado por los mirlos en el lugar experimental excede el tiempo promedio verdadero que los pájaros de este tipo emplean en el lugar natural? Pruebe las hipótesis apropiadas.

c. Calcule la diferencia entre el tiempo promedio verdadero que los mirlos emplean en el lugar natural y el tiempo promedio verdadero que los silverreyes emplean en el lugar natural y hágalo de modo que informe sobre confiabilidad y precisión.

[*Nota*: todas las medianas muestrales reportadas en el artículo parecían significativamente más pequeñas que las medias, lo que sugiere una asimetría sustancial de la distribución de población. Los autores en realidad utilizaron el procedimiento de prueba libre de distribución presentado en la sección 2 del capítulo 15.]

25. El dolor de espalda baja (DEB) es un serio problema de salud en muchos entornos industriales. El artículo "Isodynamic Evaluation of Trunk Muscles and Low-Back Pain Among Workers in a Steel Factory" (*Ergonomics*, 1995: 2107–2117) reportó los datos adjuntos sobre rango lateral de movimiento (grados) para una muestra de trabajadores sin antecedentes de dolor de espalda baja y otra muestra con antecedentes de esta dolencia.

Condición	Tamaño de muestra	Media muestral	DE muestral
Sin DEB (dolor espalda baja)	28	91.5	5.5
Con DEB	31	88.3	7.8

Calcule un intervalo de confianza de 90% para la diferencia entre el grado de movimiento lateral medio de población para las dos condiciones. ¿El intervalo sugiere que el movimiento lateral medio difiere en las dos condiciones? ¿Es diferente el mensaje si se utiliza un intervalo de confianza de 95%?

26. El artículo "The Influence of Corrosion Inhibitor and Surface Abrasion on the Failure of Aluminum-Wired Twist-On Connections" (*IEEE Trans. on Components, Hybrids, and Manuf. Tech.*, 1984: 20–25) reportó datos sobre mediciones de caída de potencial para una muestra de conectores alambrados con aluminio de aleación y otra muestra con aluminio EC. ¿Sugieren los datos adjuntos obtenidos con SAS que la caída de potencial promedio verdadera de conexiones de aleación (tipo I) es más alta que las conexiones EC (como se manifestó en el artículo)? Realice la prueba apropiada con un nivel de significación de .01. Al llegar a su conclusión, ¿qué tipo de error podría haber cometido? [*Nota*: SAS reporta el valor *P* para una prueba de dos colas.]

```
Tipo     N          Media      Desv est.     Error est
1       20   17.49900000   0.55012821   0.12301241
2       20   16.90000000   0.48998389   0.10956373

          Varianzas          T       DF     Prob>|T|
           Desigual     3.6362     37.5       0.0008
              Igual     3.6362     38.0       0.0008
```

27. La anorexia nerviosa (AN) es un trastorno psiquiátrico que lleva a la pérdida de peso importante entre las mujeres que tienen miedo a engordar. El artículo "Adipose Tissue Distribution After Weight Restoration and Weight Maintenance in Women with Anorexia Nervosa" (*Amer. J. of Clinical Nutr.*, 2009: 1132–1137) utilizó imágenes de resonancia magnética de cuerpo entero para determinar las características diferentes de tejido para una muestra de individuos con AN que se habían sometido a la recuperación del peso aguda y mantenido su peso durante un año y comparable a (al comienzo del estudio) la muestra de control. Aquí se resumen los datos sobre el tejido adiposo intermuscular (TAI; kg).

Condición	Tamaño muestral	Media muestral	Desv. est. muestral
AN	16	.52	.26
Control	8	.35	.15

Supongamos que ambas muestras fueron seleccionadas de distribuciones normales.

a. Calcule una estimación del TAI promedio real en el marco del protocolo AN descrito y hágalo de una manera que transmita información acerca de la fiabilidad y la precisión de la estimación.

b. Calcule una estimación de la diferencia entre el promedio real TAI AN y el promedio real de control del TAI y hágalo de una manera que transmita información acerca de la fiabilidad y la precisión de la estimación. ¿Qué hace que su estimación sugiera un promedio real TAI AN en relación con el promedio real de control del TAI?

28. A medida que la población envejece existe una creciente preocupación sobre lesiones relacionadas con accidentes que sufren

las personas de edad. El artículo "Age and Gender Differences in Single-Step Recovery from a Forward Fall" (*J. of Gerontology*, 1999: M44–M50) reportó sobre un experimento en el cual el ángulo de inclinación máximo, lo más lejos que un sujeto es capaz de inclinarse y aun enderezarse en un solo paso, se determinó tanto para una muestra de mujeres jóvenes (21–29 años) y una muestra de mujeres mayores (67–81 años). Las siguientes observaciones son consistentes con los datos que aparecen en el artículo:

MJ: 29, 34, 33, 27, 28, 32, 31, 34, 32, 27
MM: 18, 15, 23, 13, 12

¿Sugieren los datos que el ángulo de inclinación máximo promedio verdadero de mujeres mayores es más de 10 grados menor que el de mujeres jóvenes? Formule y pruebe las hipótesis pertinentes a nivel de significación de .10 obteniendo un valor *P*.

29. El artículo "Effect of Internal Gas Pressure on the Compression Strength of Beverage Cans and Plastic Bottles" (*J. of Testing and Evaluation*, 1993: 129–131) incluye los datos adjuntos sobre resistencia a la compresión (lb) para una muestra de latas de aluminio de 12 oz de refresco de fresas llenas y otra muestra de latas de refresco de cola llenas. ¿Sugieren los datos que la carbonatación extra de la cola da por resultado una resistencia a la compresión promedio más alta? Base su respuesta en un valor *P*. ¿Qué suposiciones son necesarias para su análisis?

Bebida	Tamaño de muestra	Media muestral	DE muestral
Fresa	15	540	21
Cola	15	554	15

30. El artículo "Flexure of Concrete Beams Reinforced with Advanced Composite Orthogrids" (*J. of Aerospace Engr.*, 1997: 7–15) dio los datos adjuntos sobre carga última (kN) de dos tipos diferentes de vigas.

Tipo	Tamaño de muestra	Media muestral	DE muestral
Vigas de fibra de vidrio	26	33.4	2.2
Vigas de fibra de carbono	26	42.8	4.3

a. Suponiendo que las distribuciones subyacentes son normales, calcule e interprete un intervalo de confianza de 99% para la diferencia entre carga promedio verdadera para las vigas de fibra de vidrio y la de vigas de fibra de carbono.
b. ¿Da el límite superior del intervalo que calculó en el inciso (a) un límite de confianza superior de 99% para la diferencia entre las dos μ? Si no, calcule tal límite. ¿Sugiere fuertemente que la carga promedio verdadera de las vigas de fibra de carbono es más grande que la de las vigas de fibra de vidrio? Explique.

31. Remítase al ejercicio 33 en la sección 7.3. El artículo citado también dio las siguientes observaciones sobre el grado de polimerización de especímenes con concentración de tiempos de viscosidad en un rango más alto:

429 430 430 431 436 437
440 441 445 446 447

a. Trace una gráfica de caja comparativa para las dos muestras y comente sobre cualquier característica interesante.
b. Calcule un intervalo de confianza de 95% para la diferencia entre el grado promedio verdadero de polimerización del rango medio y del rango alto. ¿Sugiere el intervalo que μ_1 y μ_2 pueden en realidad ser diferentes? Explique su razonamiento.

32. La enfermedad degenerativa osteoartritis afecta con mayor frecuencia a las articulaciones que soportan peso, como la rodilla. El artículo "Evidence of Mechanical Load Redistribution at the Knee Joint in the Elderly when Ascending Stairs and Ramps" (*Annals of Biomed. Engr.*, 2008: 467–476) presenta el siguiente resumen de datos sobre la duración de la postura (ms) para las muestras de ancianos y jóvenes.

Edad	Tamaño muestral	Media muestral	DE muestral
Ancianos	28	801	117
Jóvenes	16	780	72

Supongamos que ambas distribuciones de la duración de postura son normales.
a. Calcule e interprete un 99% de IC para la verdadera duración de postura media entre las personas de edad avanzada.
b. Lleve a cabo una prueba de hipótesis al nivel de significación .05 para decidir si la verdadera duración de postura promedio es mayor entre las personas de edad avanzada que entre los individuos más jóvenes.

33. El artículo "The Effects of a Low-Fat, Plant-Based Dietary Intervention on Body Weight, Metabolism, and Insulin Sensitivity in Postmenopausal Women" (*Amer. J. of Med.*, 2005: 991–997) reportó sobre los resultados de un experimento en el cual la mitad de un grupo de 64 mujeres posmenopáusicas con sobrepeso fueron asignadas al azar a un dieta vegetariana particular y la otra mitad recibió una dieta basada en las recomendaciones del National Cholesterol Education Program. La pérdida de peso media muestral de aquellas que llevaron la dieta vegetariana fue de 5.8 kg y la desviación estándar muestral fue de 3.2, en tanto que para aquellas que llevaron la dieta de control, la pérdida de peso media muestral y la desviación estándar fueron de 3.8 y 2.8, respectivamente. ¿Parece que la pérdida de peso promedio verdadera con la dieta vegetariana excede la de de la dieta de control por más de 1 kg? Realice una prueba de hipótesis apropiada a un nivel de significación de .05 basado en el cálculo de un valor *P*.

34. Considere la variable *t* agrupada

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{m} + \frac{1}{n}}}$$

la cual tiene una distribución t con $m + n - 2$ grados de libertad cuando ambas distribuciones de población son normales con $\sigma_1 = \sigma_2$ (véase la subsección de Procedimientos t agrupados para una descripción de S_p).

a. Use esta variable t para obtener una fórmula de intervalo de confianza t agrupado para $\mu_1 - \mu_2$.

b. Se seleccionó una muestra de humificadores ultrasónicos de una marca particular para la cual las observaciones de producción máxima de humedad (oz) en una cámara controlada fueron 14.0, 14.3, 12.2 y 15.1. Una muestra de una segunda marca arrojó los valores de producción 12.1, 13.6, 11.9 y 11.2 ("Multiple Comparisons of Means Using Simultaneous Confidence Intervals", *J. of Quality Technology*, 1989: 232–241). Use la fórmula de t agrupada del inciso (a) para calcular la diferencia entre producciones promedio verdaderas de las dos marcas con un intervalo de confianza de 95%.

c. Calcule la diferencia entre las dos μ utilizando el intervalo t para dos muestras discutido en esta sección y compárelo con el intervalo del inciso (b).

35. Remítase al ejercicio 34. Describa la prueba t agrupada para probar H_0: $\mu_1 - \mu_2 = \Delta_0$ cuando ambas distribuciones de población son normales con $\sigma_1 = \sigma_2$. Luego utilice este procedimiento de prueba para probar las hipótesis sugeridas en el ejercicio 33.

9.3 Análisis de datos pareados

En las secciones 9.1 y 9.2, se consideró probar en busca de una diferencia entre dos medias μ_1 y μ_2. Se hizo utilizando los resultados de una muestra aleatoria $X_1, X_2, \ldots, X_m$ de la distribución con media μ_1 y una muestra completamente independiente (de las X) $Y_1, \ldots, Y_n$ de la distribución con media μ_2. Es decir, se seleccionaron m individuos de la población 1 y n individuos diferentes de la población 2 o m individuos (u objetos experimentales) recibieron un tratamiento y otro conjunto de n individuos recibieron el otro tratamiento. En contraste, existen varias situaciones experimentales en las cuales hay sólo un conjunto de n individuos u objetos experimentales y se realizan dos observaciones de cada individuo u objeto y el resultado es un pareado natural de valores.

Ejemplo 9.8 Las trazas de metales presentes en el agua potable afectan el sabor y las concentraciones inusualmente altas plantean un riesgo para la salud. El artículo "Trace Metals of South Indian River" (*Envir. Studies*, 1982: 62–66) reporta sobre un estudio en el cual se seleccionaron seis lugares en el río (seis objetos experimentales) y se determinó la concentración de zinc (mg/L) tanto en el agua superficial como en la del fondo en cada lugar. Los seis pares de observaciones aparecen en la tabla adjunta. ¿Sugieren los datos que la concentración promedio verdadera en el agua del fondo excede la del agua de la superficie?

	Ubicación					
	1	**2**	**3**	**4**	**5**	**6**
Concentración de zinc en el agua del fondo (x)	.430	.266	.567	.531	.707	.716
Concentración de zinc en el agua de la superficie (y)	.415	.238	.390	.410	.605	.609
Diferencia	.015	.028	.177	.121	.102	.107

La figura 9.4(a) muestra una gráfica de estos datos. A primera vista, parece haber poca diferencia entre las muestras x y y. De lugar en lugar, existe mucha variación en cada muestra y parece como si cualquier diferencia entre las muestras puede ser atribuida a esta variabilidad. No obstante, cuando las observaciones están identificadas por lugar, como en la figura 9.4(b), emerge una vista diferente. En cada lugar, la concentración en el fondo excede la concentración en la superficie. Esto se confirma por el hecho de que todas las diferencias $x - y$ que aparecen en la fila inferior de la tabla son positivas. Un análisis correcto de estos datos se enfoca en estas diferencias.

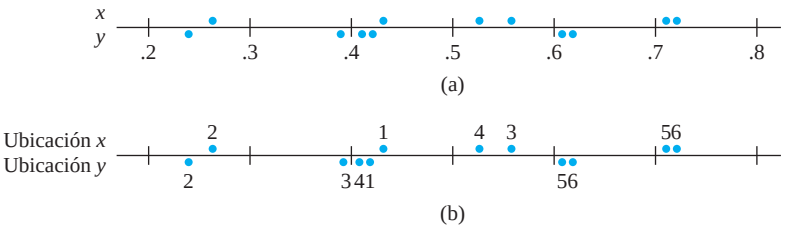


Figura 9.4 Gráfica de los datos pareados del ejemplo 9.8. (a) observaciones no identificadas por ubicación; (b) observaciones identificadas por ubicación ■

SUPOSICIONES

Los datos se componen de n pares independientemente seleccionados (X_1, Y_1), (X_2, Y_2), . . . , (X_n, Y_n), con $E(X_i) = \mu_1$ y $E(Y_i) = \mu_2$. Sean $D_1 = X_1 - Y_1$, $D_2 = X_2 - Y_2$, . . . , $D_n = X_n - Y_n$, de modo que las D_i son las diferencias dentro de los pares. En ese caso se supone que las D_i casi siempre están distribuidas con valor medio μ_D y varianza σ_D^2 (normalmente esto es una consecuencia de que las X_i y Y_i mismas están normalmente distribuidas).

De nuevo interesa probar las hipótesis con respecto a la diferencia $\mu_1 - \mu_2$. El intervalo de confianza para dos muestras t y el estadístico de prueba se obtuvo suponiendo muestras independientes y aplicando la regla $V(\overline{X} - \overline{Y}) = V(\overline{X}) + V(\overline{Y})$. Sin embargo, con datos pareados, las observaciones X y Y dentro de cada par a menudo no son independientes, de modo que $\overline{X}$ y $\overline{Y}$ no son independientes entre sí. Por consiguiente, se debe abandonar la prueba t con dos muestras y buscar un método de análisis alternativo.

Prueba t con datos pareados

Como los pares diferentes son independientes, las D_i son independientes entre sí. Sea $D = X - Y$, donde X y Y son la primera y segunda observaciones, respectivamente, dentro de un par arbitrario. Entonces la diferencia esperada es

$$\mu_D = E(X - Y) = E(X) - E(Y) = \mu_1 - \mu_2$$

(la regla de valores esperados utilizada aquí es válida aun cuando X y Y sean dependientes). Por consiguiente, cualquier hipótesis con respecto a $\mu_1 - \mu_2$ puede ser parafraseada como una hipótesis con respecto a la diferencia media μ_D. Pero como las D_i constituyen una muestra aleatoria normal (de diferencias) con media μ_D, las hipótesis con respecto a μ_D se demuestran por medio de una prueba t con una muestra. Es decir, *para probar hipótesis con respecto a* $\mu_1 - \mu_2$ *cuando los datos están pareados, se forman las diferencias* $D_1, D_2, \ldots, D_n$ *y se realiza una prueba t con una muestra* (*basada en* $n - 1$ *grados de libertad*) *de estas diferencias.*

Prueba t con datos pareados

Hipótesis nula: $H_0: \mu_D = \Delta_0$

(donde $D = X - Y$ es la diferencia entre la primera y segunda observaciones dentro de un par, y $\mu_D = \mu_1 - \mu_2$)

Valor estadístico de prueba: $t = \dfrac{\overline{d} - \Delta_0}{s_D / \sqrt{n}}$

(donde $\overline{d}$ y s_D son la media y desviación estándar muestrales, respectivamente, de las d_i)

Hipótesis alternativa	Región de rechazo para una prueba a nivel α
$H_a: \mu_D > \Delta_0$	$t \geq t_{\alpha,n-1}$
$H_a: \mu_D < \Delta_0$	$t \leq -t_{\alpha,n-1}$
$H_a: \mu_D \neq \Delta_0$	$t \geq t_{\alpha/2,n-1}$ o $t \leq -t_{\alpha/2,n-1}$

Se puede calcular un valor P como se hizo en pruebas t anteriores.

Ejemplo 9.9 Los desórdenes musculoesqueléticos del cuello y hombro son comunes entre empleados de oficina que realizan tareas repetitivas mediante pantallas de visualización. El artículo "Upper-Arm Elevation During Office Work" (*Ergonomics*, 1996: 1221–1230) reportó sobre un estudio para determinar si condiciones de trabajo más variadas habrían tenido algún impacto en el movimiento del brazo. Los datos adjuntos se obtuvieron con una muestra de $n = 16$ sujetos. Cada observación es la cantidad de tiempo, expresada como una proporción de tiempo total observado, durante el cual la elevación del brazo fue de menos de 30°. Las dos mediciones de cada sujeto se obtuvieron con una separación de 18 meses. Durante este periodo, las condiciones de trabajo cambiaron y se permitió que los sujetos realizaran una variedad más amplia de tareas. ¿Sugieren estos datos que el tiempo promedio verdadero durante el cual la elevación es menor de 30° difiere después del cambio de lo que era antes del mismo?

Sujeto	1	2	3	4	5	6	7	8
Antes	81	87	86	82	90	86	96	73
Después	78	91	78	78	84	67	92	70
Diferencia	3	−4	8	4	6	19	4	3
Sujeto	9	10	11	12	13	14	15	16
Antes	74	75	72	80	66	72	56	82
Después	58	62	70	58	66	60	65	73
Diferencia	16	13	2	22	0	12	−9	9

La figura 9.5 muestra una gráfica de probabilidad normal de las 16 diferencias; el patrón seguido por la gráfica es bastante lineal, lo que afirma la suposición de normalidad. En la figura 9.6 aparece una gráfica de caja de estas diferencias; la gráfica de caja se encuentra considerablemente a la derecha del cero, lo que sugiere que quizás $\mu_D > 0$ (observe también que 13 de las 16 diferencias son positivas y sólo dos son negativas).

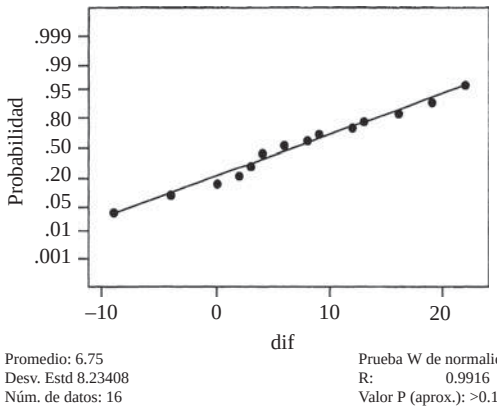


Figura 9.5 Gráfica de probabilidad normal generada por Minitab de las diferencias en el ejemplo 9.9

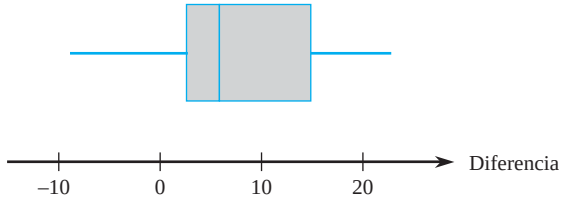


Figura 9.6 Gráfica de caja de las diferencias en el ejemplo 9.9

Pruebe ahora las hipótesis apropiadas.

1. Sea μ_D la diferencia promedio verdadera entre el tiempo de elevación antes del cambio de las condiciones de trabajo y el tiempo después del cambio.
2. H_0: $\mu_D = 0$ (no existe diferencia entre el tiempo promedio verdadero antes del cambio y el tiempo promedio verdadero después del cambio)
3. H_0: $\mu_D \neq 0$
4. $t = \dfrac{\bar{d} - 0}{s_D/\sqrt{n}} = \dfrac{\bar{d}}{s_D/\sqrt{n}}$
5. $n = 16$, $\Sigma d_i = 108$, y $\Sigma d_i^2 = 1746$, de donde $\bar{d} = 6.75$, $s_D = 8.234$, y
$$t = \frac{6.75}{8.234/\sqrt{16}} = 3.28 \approx 3.3$$
6. La tabla A.8 del apéndice muestra que el área a la derecha de 3.3 bajo la curva t con 15 grados de libertad es .002. La desigualdad de H_a implica que la prueba de dos colas es apropiada, de modo que el valor P es aproximadamente $2(.002) = .004$ (Minitab da .0051).
7. Como $.004 < .01$, la hipótesis nula puede ser rechazada a un nivel de significancia de .05 o .01. Parece que la diferencia promedio verdadera entre los tiempos es algún valor distinto de cero; es decir, el tiempo promedio verdadero después del cambio es diferente del de antes del cambio. ■

Cuando el número de pares es grande, la suposición de distribución de diferencia normal no es necesaria. El teorema del límite central valida la prueba z resultante.

Un intervalo de confianza *t* pareado

En la misma forma en que el intervalo de confianza t para una media de población única μ está basado en la variable $T = (\bar{X} - \mu)/(S/\sqrt{n})$, un intervalo de confianza t para $\mu_D\,(= \mu_1 - \mu_2)$ está basado en el hecho de que

$$T = \frac{\bar{D} - \mu_D}{S_D/\sqrt{n}}$$

tiene una distribución t con $n - 1$ grados de libertad. La manipulación de la variable t, como en deducciones previas de intervalos de confianza, produce el siguiente intervalo de confianza de $100(1 - \alpha)\%$:

El **intervalo de confianza *t* pareado para μ_D** es

$$\bar{d} \pm t_{\alpha/2,n-1} \cdot s_D/\sqrt{n}$$

Al retener el signo pertinente y al reemplazar $t_{\alpha/2}$ con t_α se obtiene un límite de confianza unilateral.

Cuando n es pequeño, la validez de este intervalo requiere que la distribución de diferencias sea por lo menos aproximadamente normal. Con n grande, el límite del teorema central garantiza que el intervalo z resultante es válido sin ninguna restricción en la distribución de diferencias.

Ejemplo 9.10 La adición de imágenes médicas computarizadas a una base de datos promete proporcionar grandes recursos para médicos. Sin embargo, existen otros métodos de obtener tal información, de modo que el tema de eficiencia de acceso tiene que ser investigado. El artículo "The Comparative Effectiveness of Conventional and Digital Image Libraries" (*J. of Audiovisual Media in Medicine*, 2001: 8–15) reportó sobre un experimento en el cual a 13 profesionistas médicos expertos en la computadora se les tomó el tiempo tanto mientras recuperaban una imagen de una biblioteca de diapositivas y mientras recuperaban la misma imagen de una base de datos de una computadora con conexión a la Web.

Sujeto	1	2	3	4	5	6	7	8	9	10	11	12	13
Diapositiva	30	35	40	25	20	30	35	62	40	51	25	42	33
Digital	25	16	15	15	10	20	7	16	15	13	11	19	19
Diferencia	5	19	25	10	10	10	28	46	25	38	14	23	14

Sea μ_D la diferencia media verdadera entre el tiempo de recuperación de diapositivas (s) y el tiempo de recuperación digital. El uso de un intervalo de confianza t pareado para estimar μ_D requiere que la distribución de diferencia sea por lo menos aproximadamente normal. La configuración lineal de los puntos en la gráfica de probabilidad normal generada por Minitab (figura 9.7) valida la suposición de normalidad. (Aparecen sólo 9 puntos debido a empates en las diferencias.)

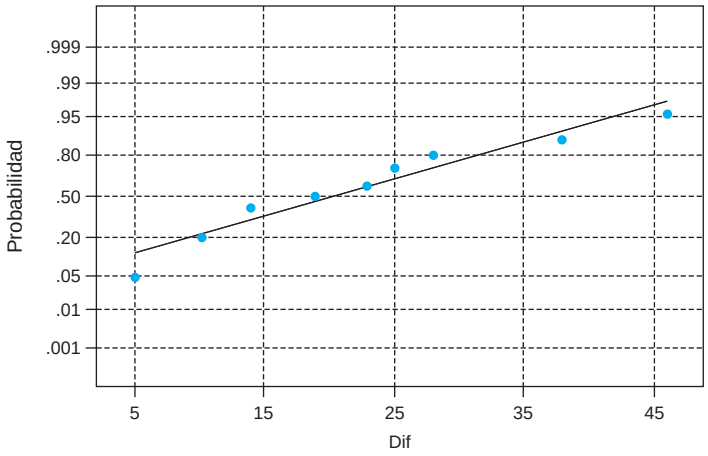


Figura 9.7 Gráfica de probabilidad normal de las diferencias en el ejemplo 9.10

Las cantidades importantes son $\Sigma d_i = 267$, $\Sigma d_i^2 = 7201$, de donde $\bar{d} = 20.5$, $s_D = 11.96$. El valor t crítico requerido para un nivel de confianza de 95% es $t_{.025,12} = 2.179$ y el intervalo de confianza de 95% es

$$\bar{d} \pm t_{\alpha/2,n-1} \cdot \frac{s_D}{\sqrt{n}} = 20.5 \pm (2.179) \cdot \frac{11.96}{\sqrt{13}} = 20.5 \pm 7.2 = (13.3, 27.7)$$

Se puede tener una plena confianza (al nivel de confianza de 95%) de que $13.3 < \mu_D < 27.7$. Este intervalo es bastante ancho, una consecuencia de que la desviación estándar muestral es relativamente grande en relación con la media muestral. Se requeriría un tamaño de muestra mucho más grande que 13 para calcular con más precisión sustancial.

Observe, sin embargo, que 0 queda muy afuera del intervalo, lo que sugiere que $\mu_D > 0$; esto se confirma con una prueba formal de hipótesis. ■

Datos pareados y procedimientos *t* con dos muestras

Considérese el uso de la prueba *t* de datos pareados con dos muestras. Los numeradores de los dos estadísticos de prueba son idénticos, puesto que $\bar{d} = \Sigma d_i/n = [\Sigma(x_i - y_i)]/n = (\Sigma x_i)/n - (\Sigma y_i)/n = \bar{x} - \bar{y}$. La diferencia entre los estadísticos se debe por completo a los denominadores. Cada estadístico de prueba se obtiene estandarizando $\overline{X} - \overline{Y}\ (=\overline{D})$. Pero en la presencia de dependencia la estandarización *t* con dos muestras es incorrecta. Para ver esto, recuérdese de la sección 5.5 que

$$V(X \pm Y) = V(X) + V(Y) \pm 2\,\mathrm{Cov}(X, Y)$$

La correlación entre X y Y es

$$\rho = \mathrm{Corr}(X, Y) = \mathrm{Cov}(X, Y)/[\sqrt{V(X)} \cdot \sqrt{V(Y)}]$$

Se desprende que

$$V(X - Y) = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$$

Aplicando esto a $\overline{X} - \overline{Y}$ se obtiene

$$V(\overline{X} - \overline{Y}) = V(\overline{D}) = V\left(\frac{1}{n}\Sigma D_i\right) = \frac{V(D_i)}{n} = \frac{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}{n}$$

La prueba *t* con dos muestras está basada en la suposición de independencia, en cuyo caso $\rho = 0$. Pero en muchos experimentos pareados, habrá una fuerte dependencia *positiva* entre X y Y (X grande asociada con Y grande), de modo que ρ será positiva y la varianza de $\overline{X} - \overline{Y}$ será más pequeña que $\sigma_1^2/n + \sigma_2^2/n$. Por lo tanto, *siempre que haya dependencia positiva dentro de los pares, el denominador del estadístico t pareado deberá ser más pequeño que para t de la prueba con muestras independientes*. Con frecuencia la *t* con dos muestras se aproximará mucho más a cero que la *t* pareada, subestimando considerablemente la significación de los datos.

Asimismo, cuando los datos están pareados, el intervalo de confianza *t* pareado normalmente será más angosto que el intervalo de confianza *t* para dos muestras (incorrecto). Esto es porque en general existe mucho menos variabilidad en las diferencias que en los valores *x* y *y*.

Experimentos pareados contra no pareados

En los ejemplos se obtuvieron datos pareados con dos observaciones del mismo sujeto (ejemplo 9.9) u objeto experimental (localización en el ejemplo 9.8). Aun cuando esto no puede hacerse, se pueden obtener datos pareados con dependencia dentro de pares emparejando individuos u objetos en relación con una o más características que se piensa influyen en las respuestas. Por ejemplo, en un experimento médico para comparar la eficacia de dos medicamentos para bajar la presión sanguínea, el presupuesto del experimentador permitiría el tratamiento de 20 pacientes. Si se seleccionan 10 al azar para tratamiento con el primer medicamento y se seleccionan otros 10 independientemente para tratamiento con el segundo medicamento, el resultado es un experimento con muestras independientes.

No obstante, el experimentador, sabiendo que la edad y el peso influyen en la presión sanguínea, podría decidir crear pares de pacientes de modo que dentro de cada uno de los 10 pares resultantes, la edad y el peso fueran aproximadamente iguales (aunque pudiera haber diferencias apreciables entre los pares). Entonces cada medicamento sería adminis-

trado a un paciente diferente dentro de cada par para un total de 10 observaciones de cada medicamento.

Sin este emparejamiento (o "bloqueo"), podría parecer que un medicamento sobrepasa el desempeño de otro simplemente porque los pacientes en una muestra pesaban menos y eran más jóvenes y por tanto más susceptibles a reducir su presión sanguínea que los pacientes más pesados y de más edad presentes en la segunda muestra. Sin embargo, hay un precio que pagar por el emparejamiento: un número de grados de libertad más pequeño para el análisis pareado, así que hay que preguntarse cuándo se debe preferir un experimento sobre el otro.

No existe una respuesta directa y precisa a esta pregunta, pero sí algunas recomendaciones útiles. Si se tiene una opción entre dos pruebas t que son válidas (y realizadas al mismo nivel de significancia α), se deberá preferir la prueba que tenga el número más grande de grados de libertad. La razón de esto es que un número más grande de grados de libertad significa una β más pequeña con cualquier valor alternativo fijo del parámetro o parámetros. Esto es, con una probabilidad de error de tipo I fija, la probabilidad de un error de tipo II se reduce al incrementarse los grados de libertad.

Sin embargo, si las unidades experimentales son bastante heterogéneas en su respuesta, será difícil detectar diferencias pequeñas pero significativas entre dos tratamientos. Esto en esencia es lo que aconteció en el conjunto de datos en el ejemplo 9.8; con ambos "tratamientos" (agua del fondo y agua superficial), existe una gran variabilidad entre lugares, lo que tiende a enmascarar diferencias en tratamientos dentro de los lugares. Si existe una alta correlación positiva dentro de unidades experimentales o sujetos, la varianza de $\overline{D} = \overline{X} - \overline{Y}$ será mucho más pequeña que la varianza no pareada. Debido a que esto reduce la varianza, será más fácil detectar una diferencia con muestras pareadas que con muestras independientes. Los pros y los contras de aparear ahora se resumen como sigue.

1. Si existe una gran heterogeneidad entre unidades experimentales y una gran correlación dentro de unidades experimentales (ρ grande positiva), entonces la pérdida de grados de libertad será compensada por la precisión incrementada asociada con el apareamiento, así que se prefiere un experimento pareado a un experimento con muestras independientes.
2. Si las unidades experimentales son relativamente homogéneas y la correlación dentro de los pares no es grande, la ganancia en precisión a causa del apareamiento será superada por la disminución de grados de libertad, así que se deberá utilizar un experimento con muestras independientes.

Desde luego, normalmente los valores de σ_1^2, σ_2^2 y ρ no serán conocidos con precisión, así que se requerirá que un investigador haga una apreciación educada sobre si se obtiene la situación 1 o la 2. En general, si el número de observaciones obtenidas es grande, entonces una pérdida de grados de libertad (p. ej., de 40 a 20) no será seria; pero si el número es pequeño, entonces la pérdida (por ejemplo, de 16 a 8) debido al apareamiento puede ser seria si no es compensada por la precisión incrementada. Consideraciones similares son válidas cuando se elige entre dos tipos de experimentos para estimar $\mu_1 - \mu_2$ con un intervalo de confianza.

EJERCICIOS Sección 9.3 (36–48)

36. Considere los datos adjuntos sobre carga de ruptura (kg/25 mm de ancho) de varias telas tanto desgastadas como no desgastadas ("The Effect of Wet Abrasive Wear on the Tensile Properties of Cotton and Polyester-Cotton Fabrics", *J. Testing and Evaluation*, 1993: 84–93). Use la prueba t pareada, como lo hicieron los autores del citado artículo, para demostrar H_0: $\mu_D = 0$ contra H_a: $\mu_D > 0$ a un nivel de significación de .01.

	Tela							
	1	2	3	4	5	6	7	8
NG	36.4	55.0	51.5	38.7	43.2	48.8	25.6	49.8
G	28.5	20.0	46.0	34.5	36.5	52.5	26.5	46.5

37. Se ha identificado cromo hexavalente como carcinógeno inhalado y como una toxina presente en el aire de interés en varios lugares diferentes. El artículo "Airborne Hexavalent Chromium in Southwestern Ontario" (*J. of Air and Waste Mgmnt. Assoc.*, 1997: 905–910) reportó los datos adjuntos tanto de concentración bajo techo como al aire libre (nanogramos/m^3) para una muestra de casas seleccionadas al azar en cierta región.

	Casa								
	1	2	3	4	5	6	7	8	9
Bajo techo	.07	.08	.09	.12	.12	.12	.13	.14	.15
Intemperie	.29	.68	.47	.54	.97	.35	.49	.84	.86

	Casa							
	10	11	12	13	14	15	16	17
Bajo techo	.15	.17	.17	.18	.18	.18	.18	.19
Intemperie	.28	.32	.32	1.55	.66	.29	.21	1.02

	Casa							
	18	19	20	21	22	23	24	25
Bajo techo	.20	.22	.22	.23	.23	.25	.26	.28
Intemperie	1.59	.90	.52	.12	.54	.88	.49	1.24

	Casa							
	26	27	28	29	30	31	32	33
Bajo techo	.28	.29	.34	.39	.40	.45	.54	.62
Intemperie	.48	.27	.37	1.26	.70	.76	.99	.36

a. Calcule un intervalo de confianza para la diferencia de media de población entre concentraciones bajo techo y a la intemperie utilizando un nivel de confianza de 95% e interprete el intervalo resultante.

b. Si la 34ª casa fuera seleccionada al azar de la población, ¿entre qué valores pronosticaría que quede la diferencia de concentraciones?

38. Se sacaron especímenes de concreto con proporciones variables de altura a diámetro de varias posiciones cortadas del cilindro original tanto de una mezcla de concreto de resistencia normal como de una mezcla de alta resistencia. Se determinó el esfuerzo pico (MPa) de cada mezcla y se obtuvieron los siguientes datos ("Effect of Length on Compressive Strain Softening of Concrete", *J. of Engr. Mechanics*, 1997: 25–35):

	Condición de prueba				
	1	2	3	4	5
Normal	42.8	55.6	49.0	48.7	44.1
Alta	90.9	93.1	86.3	90.3	88.5

	Condición de prueba				
	6	7	8	9	10
Normal	55.4	50.1	45.7	51.4	43.1
Alta	88.1	93.2	90.8	90.1	92.6

	Condición de prueba				
	11	12	13	14	15
Normal	46.8	46.7	47.7	45.8	45.4
Alta	88.2	88.6	91.0	90.0	90.1

a. Construya una gráfica de caja comparativa de esfuerzos pico para los dos tipos de concreto y comente sobre cualquier característica interesante.

b. Estime la diferencia entre esfuerzos pico promedio verdaderos de los dos tipos de concreto en una forma que transmita información sobre precisión y confiabilidad. Asegúrese de verificar la factibilidad de cualquier suposición requerida en su análisis. ¿Parece factible que los esfuerzos pico promedio verdaderos para los dos tipos de concreto sean idénticos? ¿Por qué sí o por qué no?

39. Científicos e ingenieros con frecuencia desean comparar dos técnicas diferentes de medir o determinar el valor de una variable. En tales situaciones, el interés se concentra en probar si la diferencia media en las mediciones es cero. El artículo "Evaluation of the Deuterium Dilution Technique Against the Test Weighing Procedure for the Determination of Breast Milk Intake" (*Amer. J. of Clinical Nutr.*, 1983: 996–1003) reporta los datos adjuntos sobre la cantidad de leche ingerida por cada uno de 14 infantes seleccionados al azar.

	Infante				
	1	2	3	4	5
Método isotrópico	1509	1418	1561	1556	2169
Método ponderado de prueba	1498	1254	1336	1565	2000
Diferencia	11	164	225	−9	169

	Infante				
	6	7	8	9	10
Método isotrópico	1760	1098	1198	1479	1281
Método ponderado de prueba	1318	1410	1129	1342	1124
Diferencia	442	−312	69	137	157

	Infante			
	11	12	13	14
Método isotrópico	1414	1954	2174	2058
Método ponderado de prueba	1468	1604	1722	1518
Diferencia	−54	350	452	540

a. ¿Es factible que la distribución de población de las diferencias sea normal?

b. ¿Parece que la diferencia promedio verdadera entre valores de ingesta medidos con los dos métodos es algún valor diferente de cero? Determine el valor *P* de la prueba y utilícelo para llegar a una conclusión a nivel de significación de .05.

40. La lactancia estimula una pérdida temporal de masa ósea para proporcionar cantidades de calcio adecuadas para la producción de leche. El artículo "Bone Mass Is Recovered from Lactation to Postweaning in Adolescent Mothers with Low Calcium Intakes" (*Amer. J. of Clinical Nutr.*, 2004; 1322-1326) dio los siguientes datos sobre contenido total de minerales en los huesos del cuerpo (TBBMC, por sus siglas en inglés) (g) para una muestra tanto durante la lactancia (L) como en el periodo de posdestete (P).

	Sujeto									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
L	1928	2549	2825	1924	1628	2175	2114	2621	1843	2541
P	2126	2885	2895	1942	1750	2184	2164	2626	2006	2627

a. ¿Sugieren los datos que el contenido total de minerales en los huesos del cuerpo durante el posdestete excede el de la lactancia por más de 25 g? Formule y pruebe las hipótesis apropiadas utilizando un nivel de significación de .05 [*Nota*: la gráfica de probabilidad normal apropiada muestra algo de curvatura pero no suficiente para sembrar dudas sustanciales sobre una suposición de normalidad.]

b. Calcule un límite de confianza superior utilizando un nivel de confianza de 95% para la diferencia promedio verdadera entre TBBMC durante el posdestete y durante la lactancia.

c. ¿Conduce el uso (incorrecto) de la prueba *t* con dos muestras para demostrar las hipótesis sugeridas en (a) a la misma conclusión a la que se llegó allí? Explique.

41. Los fármacos antipsicóticos son ampliamente prescritos para condiciones como la esquizofrenia y enfermedad bipolar. El artículo "Cardiometabolic Risk of Second-Generation Antipsychotic Medications During First-Time Use in Children and Adolescents" (*J. of the Amer. Med. Assoc.*, 2009) informó sobre la composición corporal y cambios en el metabolismo de las personas que habían tomado varios medicamentos antipsicóticos por periodos cortos de tiempo.

a. La muestra de 41 individuos que habían tomado aripiprazol tuvo una variación media del colesterol total (mg/dL) de 3.75, y el error estándar estimado $s_D/\sqrt{n}$ fue de 3.878. Calcule un intervalo de confianza con un nivel de confianza de aproximadamente el 95% para el verdadero aumento promedio en el colesterol total en estas circunstancias (el artículo citado incluyó este IC).

b. El artículo también informaba que en una muestra de 36 individuos que habían tomado la quetiapina, la media de la muestra del nivel de colesterol cambia y el error estándar estimado fue 9.05 y 4.256, respectivamente. Hacer cualquier suposición necesaria acerca de la distribución de los cambios en el nivel de colesterol, ¿influye en la elección de un nivel de significación para su conclusión acerca de cierto aumento promedio en el nivel de colesterol? Explique. [*Nota*: el artículo incluye un valor *P*.]

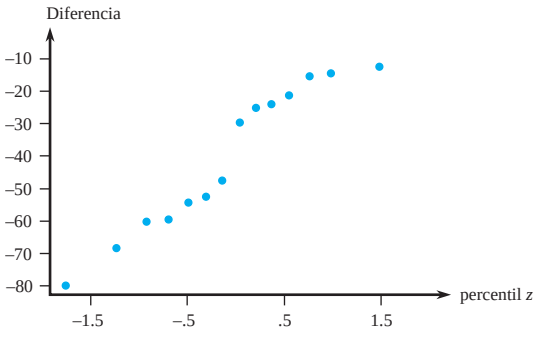
c. Para la muestra de 45 individuos que habían tomado la olanzapina, el artículo reportó (7.38, 9.69) como un IC del 95% para la ganancia real de peso promedio (kg). ¿Qué es un IC del 99%?

42. Se ha estimado que entre 1945 y 1971 nacieron 2 millones de niños de madres tratadas con dietilestibrestrol (DES, por sus siglas en inglés) un estrógeno no esteroidal recomendado para el mantenimiento del embarazo. La FDA (Federal Drug Administration) vetó este medicamento en 1971 porque investigaciones indicaron que había una conexión con la incidencia de cáncer cervical. El artículo "Effects of Prenatal Exposure to Diethylstilbestrol (DES) on Hemispheric Laterality and Spatial Ability in Human Males" (*Hormones and Behavior*, 1992: 62–75) discutió un estudio en el cual 10 varones expuestos a DES y sus hermanos no expuestos fueron sometidos a varias pruebas. Éstos son los datos sobre los resultados de una prueba de habilidad espacial: $\bar{x} = 12.6$ (expuestos), $\bar{y} = 13.7$, y error estándar de la diferencia media $= .5$. Pruebe a un nivel de .05 para ver si la exposición tiene que ver con la habilidad espacial reducida mediante la obtención del valor *P*.

43. La enfermedad de Cushing se caracteriza por debilidad muscular por una disfunción de la suprarrenal o pituitaria. Para administrar un tratamiento eficaz, es importante detectar la enfermedad de Cushing en la niñez tan pronto como sea posible. La edad al inicio de los síntomas (meses) y la edad en el momento del diagnóstico en 15 niños que padecen la enfermedad aparecieron en el artículo "Treatment of Cushing's Disease in Childhood and Adolescence by Transphenoidal Microadenomectomy" (*New Engl. J. of Med.*, 1984: 889). A continuación se dan los valores de las diferencias de edades al principio de los síntomas y la edad en el momento del diagnóstico:

$-24 \quad -12 \quad -55 \quad -15 \quad -30 \quad -60 \quad -14 \quad -21$
$-48 \quad -12 \quad -25 \quad -53 \quad -61 \quad -69 \quad -80$

a. ¿Siembra una fuerte duda la gráfica de probabilidad normal adjunta sobre la normalidad aproximada de la distribución de diferencias de población?



b. Calcule un límite de confianza de 95% inferior para la diferencia media de población e interprete el límite resultante.

c. Suponga que ya se habían calculado las diferencias (edad al momento del diagnóstico) − (edad al inicio de los síntomas). ¿Cuál sería un límite de confianza superior de 95% para la diferencia de la media de población correspondiente?

44. Refiérase al ejercicio anterior.

a. Con mucho, la hipótesis de mayor frecuencia fue nula cuando los datos estaban pareados es H_0: $\mu_D = 0$. ¿Es una hipótesis razonable en este contexto? Explique.

b. Lleve a cabo una prueba de hipótesis para decidir si existen pruebas convincentes para concluir que el diagnóstico se produce en promedio más de 25 meses después de la aparición de los síntomas.

45. Torsión durante la rotación externa de la cadera (RE) y la extensión pueden ser responsables de ciertos tipos de lesiones en jugadores de golf y otros atletas. El artículo "Hip Rotational Velocities During the Full Golf Swing" (*J. of Sports Science and Medicine*, 2009: 296–299) informó sobre un estudio en que el pico de velocidad de RE y el pico de velocidad (rotación interna) de IR (ambas en grado.segundo^{-1}) fue determinado en una muestra de 15 golfistas femeninas colegiales durante sus giros. Los siguientes datos son suministrados por los autores del artículo

Golfista	RE	RI	Diferencia	Percentil z
1	−130.6	−98.9	−31.7	−1.28
2	−125.1	−115.9	−9.2	−0.97
3	−51.7	−161.6	109.9	0.34
4	−179.7	−196.9	17.2	−0.73
5	−130.5	−170.7	40.2	−0.34
6	−101.0	−274.9	173.9	0.97
7	−24.4	−275.0	250.6	1.83
8	−231.1	−275.7	44.6	−0.17
9	−186.8	−214.6	27.8	−0.52
10	−58.5	−117.8	59.3	0.00
11	−219.3	−326.7	107.4	0.17
12	−113.1	−272.9	159.8	0.73
13	−244.3	−429.1	184.8	1.28
14	−184.4	−140.6	−43.8	−1.83
15	−199.2	−345.6	146.4	0.52

a. ¿Es factible que las diferencias provengan de una población distribuida normalmente?

b. El artículo informó que Media (± DE) = −145.3(68.0) para la velocidad de RE e = −227.8(96.6) para la velocidad del RI. Basándose sólo en esta información, ¿podría realizarse una prueba de hipótesis acerca de la diferencia entre la velocidad real media del RI y la velocidad real media de RE? Explique.

c. El artículo afirmaba que "el pico principal de velocidad del RI de la cadera fue significativamente mayor que la velocidad de arrastre RE de la cadera ($p = .003$, valor $t = 3.65$)". (La redacción sugiere que se utilizó una prueba de cola superior.) De hecho, ¿éste es en realidad el caso? [*Nota*: "$p = .033$" en la tabla 2 de este artículo es errónea.]

46. El ejemplo 7.11 aportó datos sobre el módulo de elasticidad obtenido 1 minuto después de cargar con una configuración de especímenes de madera. El artículo citado también aportó los valores del módulo de elasticidad obtenidos 4 semanas después de cargar los mismos especímenes de madera. A continuación se presentan los datos.

Observación	1 minuto	4 semanas	Diferencia
1	10,490	9,110	1380
2	16,620	13,250	3370
3	17,300	14,720	2580
4	15,480	12,740	2740
5	12,970	10,120	2850
6	17,260	14,570	2690
7	13,400	11,220	2180
8	13,900	11,100	2800
9	13,630	11,420	2210
10	13,260	10,910	2350
11	14,370	12,110	2260
12	11,700	8,620	3080
13	15,470	12,590	2880
14	17,840	15,090	2750
15	14,070	10,550	3520
16	14,760	12,230	2530

Calcule e interprete un límite de confianza superior para la diferencia promedio verdadera entre el módulo después de 1 minuto y el módulo después de 4 semanas; primero compruebe la factibilidad de cualquier suposición necesaria.

47. El artículo "Slender High-Strength RC Columns Under Eccentric Compression" (*Magazine of Concrete Res.*, 2005: 361–370) dio los datos adjuntos sobre resistencia de cilindros (MPa) de varios tipos de columnas curadas tanto en condiciones húmedas como en condiciones secas en el laboratorio.

	Tipo					
	1	**2**	**3**	**4**	**5**	**6**
H:	82.6	87.1	89.5	88.8	94.3	80.0
CS:	86.9	87.3	92.0	89.3	91.4	85.9
	7	**8**	**9**	**10**	**11**	**12**
H:	86.7	92.5	97.8	90.4	94.6	91.6
CS:	89.4	91.8	94.3	92.0	93.1	91.3

a. Estime la diferencia en la resistencia promedio verdadera en las dos condiciones secas en una forma que dé información sobre confiabilidad y precisión e interprete la estimación. ¿Qué sugiere la estimación sobre cómo se compara la resistencia promedio verdadera en condiciones húmedas y en condiciones secas en el laboratorio?

b. Verifique la plausibilidad de cualquier suposición que fundamenten su análisis de (a).

48. Construya un conjunto de datos pareados para el cual $t = \infty$, de modo que los datos sean altamente significativos cuando se utilice el análisis correcto, aunque t para la prueba t con dos muestras esté bastante cerca de cero, de tal suerte que el análisis incorrecto dé un resultado insignificante.

9.4 Inferencias sobre una diferencia entre proporciones de población

Después de presentar métodos para comparar las medias de dos poblaciones diferentes, ahora se presta atención a la comparación de dos proporciones de población. Un individuo u objeto se considera como éxito S si él/ella/ello posee alguna característica de interés (alguien que se graduó de una universidad, un refrigerador con hacedor de cubos de hielo, etc.). Sea

$$p_1 = \text{la proporción de éxitos } (S) \text{ en la población } \#\,1$$
$$p_2 = \text{la proporción de éxitos } (S) \text{ en la población } \#\,2$$

Alternativamente, $p_1(p_2)$ pueden ser consideradas como la probabilidad de que un individuo u objeto seleccionado al azar de la primera (segunda) población sea un éxito.

Supóngase que se selecciona un tamaño de muestra m de la primera población e independientemente, se selecciona una muestra de tamaño n de la segunda. Sea X el número de éxitos (S) en la primera muestra y Y el número de éxitos (S) en la segunda. La independencia de las dos muestras implica que X y Y son independientes. Siempre que los dos tamaños de muestras sean mucho más pequeños que los tamaños de población correspondientes, se puede considerar que las distribuciones de X y Y son binomiales. El estimador natural de $p_1 - p_2$, la diferencia en las proporciones de la población, es la diferencia correspondiente en las proporciones muestrales $X/m - Y/n$.

PROPOSICIÓN

Sean $\hat{p}_1 = X/m$ y $\hat{p}_2 = Y/n$, donde $X \sim \text{Bin}(m, p_1)$ y $Y \sim \text{Bin}(n, p_2)$ con X y Y variables independientes. Entonces

$$E(\hat{p}_1 - \hat{p}_2) = p_1 - p_2$$

de modo que $\hat{p}_1 - \hat{p}_2$ sea un estimador insesgado de $p_1 - p_2$, y

$$V(\hat{p}_1 - \hat{p}_2) = \frac{p_1q_1}{m} + \frac{p_2q_2}{n} \quad (\text{donde } q_i = 1 - p_i) \tag{9.3}$$

Demostración Como $E(X) = mp_1$ y $E(Y) = np_2$,

$$E\left(\frac{X}{m} - \frac{Y}{n}\right) = \frac{1}{m}E(X) - \frac{1}{n}E(Y) = \frac{1}{m}mp_1 - \frac{1}{n}np_2 = p_1 - p_2$$

Como $V(X) = mp_1q_1$, $V(Y) = np_2q_2$, y X y Y son independientes,

$$V\left(\frac{X}{m} - \frac{Y}{n}\right) = V\left(\frac{X}{m}\right) + V\left(\frac{Y}{n}\right) = \frac{1}{m^2}V(X) + \frac{1}{n^2}V(Y) = \frac{p_1q_1}{m} + \frac{p_2q_2}{n}$$

■

Primero se abordarán situaciones en las que tanto m como n son grandes. Entonces como las distribuciones de $\hat{p}_1$ y $\hat{p}_2$ son aproximadamente normales, la distribución del estimador $\hat{p}_1 - \hat{p}_2$ también es normal en forma aproximada. Al estandarizar $\hat{p}_1 - \hat{p}_2$ se obtiene una variable Z cuya distribución es aproximadamente normal estándar.

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\dfrac{p_1q_1}{m} + \dfrac{p_2q_2}{n}}}$$

Procedimiento de prueba con muestra grande

La hipótesis nula más general que un investigador podría considerar sería de la forma H_0: $p_1 - p_2 = \Delta_0$. Aunque para medias de población el caso $\Delta_0 \neq 0$ no presentó dificultades, para proporciones de población los casos $\Delta_0 = 0$ y $\Delta_0 \neq 0$ deben ser considerados por separado. Como la mayoría de los problemas reales de esta clase implican $\Delta_0 = 0$ (es decir, la hipótesis nula $p_1 = p_2$), se abordará este caso. Cuando H_0: $p_1 - p_2 = 0$ es verdadera, sea p el valor común de p_1 y p_2 (y del mismo modo para q). Entonces la variable estandarizada

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{pq\left(\frac{1}{m} + \frac{1}{n}\right)}} \tag{9.4}$$

tiene aproximadamente una distribución estándar normal cuando H_0 es verdadera. Sin embargo, esta Z no sirve como estadístico de prueba porque el valor de p es desconocido, H_0 afirma sólo que existe un valor común de p, pero no dice cuál es ese valor. Al reemplazar p y q en (9.4) por estimadores apropiados se obtiene un estadístico de prueba.

Suponiendo que $p_1 = p_2 = p$, en lugar de muestras separadas de tamaño m y n de dos poblaciones diferentes (dos distribuciones binomiales distintas), en realidad se tiene una sola muestra de tamaño $m + n$ de una población con proporción p. El número total de individuos en esta muestra combinada que tiene la característica de interés es $X + Y$. El estimador natural de p es entonces

$$\hat{p} = \frac{X + Y}{m + n} = \frac{m}{m + n} \cdot \hat{p}_1 + \frac{n}{m + n} \cdot \hat{p}_2 \tag{9.5}$$

La segunda expresión para $\hat{p}$ muestra que en realidad es un promedio ponderado de los estimadores $\hat{p}_1$ y $\hat{p}_2$ obtenidos con las dos muestras. Si se utiliza $\hat{p}$ y $\hat{q} = 1 - \hat{p}$ en lugar de p y q en (9.4) se obtiene un estadístico de prueba cuya distribución es aproximadamente normal estándar cuando H_0 es verdadera.

Hipótesis nula: H_0: $p_1 - p_2 = 0$

Valor estadístico de prueba (muestras grandes): $z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{m} + \frac{1}{n}\right)}}$

Hipótesis alternativa	Región de rechazo para una prueba a nivel α aproximado
H_a: $p_1 - p_2 > 0$	$z \geq z_\alpha$
H_a: $p_1 - p_2 < 0$	$z \leq -z_\alpha$
H_a: $p_1 - p_2 \neq 0$	$z \geq z_{\alpha/2}$ o $z \leq -z_{\alpha/2}$

Se calcula un valor P del mismo modo que para pruebas z previas.
La prueba de seguridad se puede usar siempre y cuando $m\hat{p}_1$, $m\hat{q}_1$, $n\hat{p}_2$ y $n\hat{q}_2$ sean todos al menos 10.

Ejemplo 9.11 El artículo "Aspirin Use and Survival After Diagnosis of Colorectal Cancer" (*J. of the Amer. Med. Assoc.*, 2009: 649–658) informó que de 549 participantes del estudio que utilizan regularmente aspirina después de ser diagnosticados con cáncer colorrectal, había 81 muertes por cáncer colorrectal específico, mientras que de entre 730 individuos diagnosticados de manera similar que no hicieron posteriormente uso de la aspirina, había 141

muertes por cáncer colorrectal específico. ¿Estos datos sugieren que el uso regular de aspirina después del diagnóstico, reducirá la tasa de incidencia específica de muertes por cáncer colorrectal? Probemos la hipótesis apropiada con un nivel de significación de .05.

El parámetro de interés es la diferencia $p_1 - p_2$, donde p_1 es la verdadera proporción de las muertes de aquellos que utilizan regularmente aspirina y p_2 es la verdadera proporción de muertes para los que no hicieron uso de la aspirina. El uso de la aspirina es beneficiosa si $p_1 < p_2$, que corresponde a una diferencia negativa entre las dos proporciones. Las hipótesis relevantes son por lo tanto

$$H_0: p_1 - p_2 = 0 \quad \text{contra} \quad H_a: p_1 - p_2 < 0$$

los parámetros estimados son $\hat{p}_1 = 81/549 = .1475$, $\hat{p}_2 = 141/730 = .1932$, y $\hat{p} = (81 + 141)/(549 + 730) = .1736$. Una prueba z es apropiada en este caso porque todos $m\hat{p}_1$, $m\hat{q}_1$, $n\hat{p}_2$ y $n\hat{q}_2$ son por lo menos 10. El valor de la prueba estadística resultante es

$$z = \frac{.1475 - .1932}{\sqrt{(.1736)(.8264)\left(\frac{1}{549} + \frac{1}{730}\right)}} = \frac{-.0457}{.021397} = -2.14$$

El correspondiente valor P para una prueba z de cola inferior es $\Phi(-2.14) = .0162$. Ya que $.0162 \leq .05$, la hipótesis nula puede ser rechazada al nivel de significación .05. Así que cualquiera que acepte este nivel de significación puede estar convencido de que el uso de aspirina en estas circunstancias es beneficioso. Sin embargo, alguien en busca de más evidencia convincente puede seleccionar un nivel de significancia de .01 y luego no ser convencido. ■

Probabilidades de error de tipo II y tamaños de muestra

En este caso la determinación de β es un poco más tediosa de lo que fue para otras pruebas con muestra grande. La razón es que el denominador de Z es una estimación de la desviación estándar de $\hat{p} - \hat{p}_2$, suponiendo que $p_1 = p_2 = p$. Cuando H_0 es falsa, $\hat{p}_1 - \hat{p}_2$ debe ser reestandarizada por medio de

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1 q_1}{m} + \frac{p_2 q_2}{n}} \tag{9.6}$$

La forma de σ implica que β no es una función de sólo $p_1 - p_2$, de modo que se la denota como $\beta(p_1, p_2)$.

Hipótesis alternativa	$\beta(p_1, p_2)$
$H_a: \quad p_1 - p_2 > 0$	$\Phi\left[\dfrac{z_\alpha\sqrt{\bar{p}\,\bar{q}\left(\frac{1}{m} + \frac{1}{n}\right)} - (p_1 - p_2)}{\sigma}\right]$
$H_a: \quad p_1 - p_2 < 0$	$1 - \Phi\left[\dfrac{-z_\alpha\sqrt{\bar{p}\,\bar{q}\left(\frac{1}{m} + \frac{1}{n}\right)} - (p_1 - p_2)}{\sigma}\right]$

Hipótesis alternativa	$\beta(p_1, p_2)$
H_a: $p_1 - p_2 \neq 0$	$\Phi\left[\dfrac{z_{\alpha/2}\sqrt{\bar{p}\,\bar{q}\left(\frac{1}{m}+\frac{1}{n}\right)} - (p_1 - p_2)}{\sigma}\right] - \Phi\left[\dfrac{-z_{\alpha/2}\sqrt{\bar{p}\,\bar{q}\left(\frac{1}{m}+\frac{1}{n}\right)} - (p_1 - p_2)}{\sigma}\right]$

donde $\bar{p} = (mp_1 + np_2)/(m + n)$, $\bar{q} = (mq_1 + nq_2)/(m + n)$ y σ está dada por (9.6).

Demostración Para la prueba de cola superior (H_a: $p_1 - p_2 > 0$),

$$\beta(p_1, p_2) = P\left[\hat{p}_1 - \hat{p}_2 < z_\alpha\sqrt{\hat{p}\hat{q}\left(\frac{1}{m} + \frac{1}{n}\right)}\right]$$

$$= P\left[\frac{(\hat{p}_1 - \hat{p}_2 - (p_1 - p_2))}{\sigma} < \frac{z_a\sqrt{\hat{p}\hat{q}\left(\frac{1}{m} + \frac{1}{n}\right)} - (p_1 - p_2)}{\sigma}\right]$$

Cuando m y n son grandes,

$$\hat{p} = (m\hat{p}_1 + n\hat{p}_2)/(m + n) \approx (mp_1 + np_2)/(m + n) = \bar{p}$$

y $\hat{q} \approx \bar{q}$, la cual da la expresión previa (aproximada) para $\beta(p_1, p_2)$. ■

Alternativamente, para p_1 especificada, p_2 con $p_1 - p_2 = d$, se pueden determinar los tamaños de muestra necesarios para obtener $\beta(p_1, p_2) = \beta$. Por ejemplo, para la prueba de cola superior, se iguala $-z_\beta$ al argumento de $\Phi(\cdot)$(es decir, lo que está entre paréntesis) en el recuadro siguiente. Si $m = n$, existe una expresión simple para el valor común.

En el caso $m = n$, la prueba a nivel α tiene una probabilidad β de error de tipo II con valores alternativos de p_1, p_2 con $p_1 - p_2 = d$ cuando

$$n = \frac{\left[z_\alpha\sqrt{(p_1 + p_2)(q_1 + q_2)/2}\right] + z_\beta\sqrt{p_1q_1 + p_2q_2}\right]^2}{d^2} \tag{9.7}$$

para una prueba de cola superior o inferior, con $\alpha/2$ reemplazando a α para una prueba de dos colas.

Ejemplo 9.12 Una de las aplicaciones verdaderamente impresionantes de la estadística ocurrió en conexión con el diseño de experimento y análisis de la vacuna Salk contra la polio en 1954. Una parte del experimento se enfocó en la eficacia de la vacuna en el combate de la polio paralítica. Debido a que se pensó que sin un grupo de control de niños, no habría una base sólida para evaluar la vacuna, se decidió administrar la vacuna a un grupo y una inyección placebo (visualmente indistinguible de la vacuna pero que no tiene ningún efecto) a un grupo de control. Por razones éticas y también porque se pensaba que el conocimiento de la administración de la vacuna podría afectar el tratamiento y diagnóstico, el experimento

se llevó a cabo de una manera **doblemente a ciegas.** Es decir, ninguno de los individuos que recibieron inyecciones ni los que la administraron en realidad sabían quién estaba recibiendo la vacuna y quién estaba recibiendo el placebo (las muestras fueron numéricamente codificadas); recuerde que en ese momento no estaba del todo claro si la vacuna era benéfica.

Sean p_1 y p_2 las probabilidades de que un niño contraiga polio paralítica en las condiciones de control y tratamiento, respectivamente. El objetivo era probar H_0: $p_1 - p_2 = 0$ contra H_a: $p_1 - p_2 > 0$ (la alternativa afirma que es menos probable que un niño vacunado contraiga polio que un niño no vacunado). Suponiendo que el valor verdadero de p_1 es .0003 (un coeficiente de incidencia de 30 por cada 100,000), la vacuna sería una mejora significativa si el coeficiente de incidencia se reducía a la mitad, es decir, $p_2 = .00015$. Con una prueba a un nivel $\alpha = .05$, sería entonces razonable requerir tamaños de muestra con los cuales $\beta = .1$ cuando $p_1 = .0003$ y $p_2 = .00015$. Si se suponen tamaños de muestra iguales, el requerido n se obtiene con (9.7) como

$$n = \frac{\left[1.645\sqrt{(.5)(.00045)(1.99955)} + 1.28\sqrt{(.00015)(.99985) + (.0003)(.9997)}\right]^2}{(.0003 - .00015)^2}$$

$$= [(.0349 + .0271)/.00015]^2 \approx 171{,}000$$

Los datos reales para este experimento son los siguientes. Se utilizaron tamaños de muestra de aproximadamente 200,000. El lector puede verificar con facilidad que $z = 6.43$, un valor muy significativo. ¡La vacuna fue un rotundo éxito!

Placebo: $m = 201{,}229$, $x =$ número de casos de polio paralítica $= 110$

Vacuna: $n = 200{,}745$, $y = 33$ ■

Intervalo de confianza con muestra grande

Como con las medias, muchos problemas de dos muestras implican el objetivo de comparación mediante pruebas de hipótesis, pero en ocasiones una estimación de intervalo para $p_1 - p_2$ es apropiada. Tanto $\hat{p}_1 = X/m$ y $\hat{p}_2 = Y/n$ tienen distribuciones aproximadamente normales cuando tanto m como n son grandes. Si se identifica θ con $p_1 - p_2$, entonces $\hat{\theta} = \hat{p}_1 - \hat{p}_2$ satisface las condiciones necesarias a fin de obtener un intervalo de confianza para muestra grande. En particular, la desviación estándar estimada de $\hat{\theta}$ es $\sqrt{(\hat{p}_1\hat{q}_1/m) + (\hat{p}_2\hat{q}_2/n)}$. El intervalo general $\hat{\theta} \pm z_{\alpha/2} \cdot \hat{\sigma}_{\hat{\theta}}$ de $100(1 - \alpha)\%$ toma la forma siguiente.

Un IC para $p_1 - p_2$ con un nivel de confianza aproximadamente $100(1 - \alpha)\%$ es

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{m} + \frac{\hat{p}_2\hat{q}_2}{n}}$$

Este intervalo puede usarse con seguridad siempre y cuando $m\hat{p}_1$, $m\hat{q}_1$, $n\hat{p}_2$ y $n\hat{q}_2$ sean todos al menos 10.

Obsérvese que la desviación estándar estimada de $\hat{p}_1 - \hat{p}_2$ (la expresión de la raíz cuadrada) es diferente aquí de lo que fue para probar hipótesis cuando $\Delta_0 = 0$.

Investigaciones recientes han demostrado que el nivel de confianza para el intervalo de confianza tradicional que se acaba de dar en ocasiones se desvía sustancialmente del nivel nominal (el nivel que se piensa se va a obtener cuando se utiliza un valor crítico z particular, p. ej., 95% cuando $z_{\alpha/2} = 1.96$). Se dice que la mejora sugerida es agregar un éxito y una falla a cada una de las dos muestras y luego reemplazar las $\hat{p}$ y $\hat{q}$ en la fórmula

anterior por los $\tilde{p}$ y $\tilde{q}$ donde $\tilde{p}_1 = (x + 1)/(m + 2)$, etc. Este intervalo modificado también puede ser utilizado cuando los tamaños de muestra son bastante pequeños.

Ejemplo 9.13 Los autores del artículo "Adjuvant Radiotherapy and Chemotherapy in Node-Positive Premenopausal Women with Breast Cancer" (*New Engl. J. of Med.*, 1997: 956–962) reportaron los resultados de un experimento diseñado para comparar el tratamiento de pacientes con cáncer con sólo quimioterapia con un tratamiento combinado de quimioterapia y radiación. De las 154 pacientes que recibieron el tratamiento de sólo quimioterapia, 76 sobrevivieron por lo menos 15 años, en tanto que 98 de las 164 pacientes que recibieron el tratamiento híbrido sobrevivieron por lo menos ese número de años. Con p_1 denotando la proporción de todas las mujeres que, cuando fueron tratadas con sólo quimioterapia, sobreviven por lo menos 15 años y p_2 denotando la proporción análoga para el tratamiento híbrido, $\hat{p}_1 = 76/154 = .494$ y $98/164 = .598$. Un intervalo de confianza para la diferencia entre proporciones basadas en la fórmula tradicional con un nivel de confianza de aproximadamente 99% es

$$.494 - .598 \pm (2.58)\sqrt{\frac{(.494)(.506)}{154} + \frac{(.598)(.402)}{164}} = -.104 \pm .143$$
$$= (-.247, .039)$$

Al nivel de confianza de 99%, es factible que $-.247 < p_1 - p_2 < .039$. Este intervalo es ancho de manera razonable, un reflejo del hecho de que los tamaños de muestra no son terriblemente grandes para este tipo de intervalo. Obsérvese que 0 es uno de los valores factibles de $p_1 - p_2$, lo que sugiere que ningún tratamiento puede ser juzgado superior al otro. Con $\tilde{p}_1 = 77/156 = .494$, $\tilde{q}_1 = 79/156 = .506$, $\tilde{p}_2 = .596$, $\tilde{q}_2 = .404$, con base en tamaños de muestra de 156 y 166, respectivamente, el intervalo "mejorado" aquí es idéntico al intervalo anterior. ■

Inferencias basadas en muestras pequeñas

En ocasiones una inferencia con respecto a $p_1 - p_2$ es posible que tenga que basarse en muestras donde por lo menos un tamaño de muestra es pequeño. Los métodos apropiados para tales situaciones no son directos como aquellos para muestras grandes y existe más controversia entre los estadísticos en cuanto a los procedimientos recomendados. Una prueba utilizada con frecuencia, llamada prueba de Fisher-Irwin, se basa en la distribución hipergeométrica. Su amigable estadístico vecino puede ser consultado para más información.

EJERCICIOS Sección 9.4 (49–58)

49. ¿Es menos probable que alguien que cambia de marca por cuestiones financieras permanezca leal que alguien que cambia sin pensar en cuestiones financieras? Sean p_1 y p_2 las proporciones verdaderas de los que cambian a cierta marca con o sin pensar en cuestiones financieras, respectivamente, que después repiten una compra. Pruebe H_0: $p_1 - p_2 = 0$ contra H_a: $p_1 - p_2 < 0$ con $\alpha = .01$ y los siguientes datos:

$m = 200$ número de éxitos $= 30$
$n = 600$ número de éxitos $= 180$

(Datos similares aparecen en "Impact of Deals and Deal Retraction on Brand Switching", *J. of Marketing*, 1980: 62–70.)

50. Los recientes incidentes de contaminación de los alimentos han causado gran preocupación entre los consumidores. El artículo "How Safe Is That Chicken?" (*Consumer Reports*, enero de 2010: 19–23) informó que 35 de los 80 pollos seleccionados al azar de la marca Perdue dieron positivo, ya sea para campylobacter o salmonella (o ambos), las principales causas bacterianas de enfermedades transmitidas por los alimentos, mientras que 66 de 80 pollos de la marca Tyson dieron positivo.

a. ¿Parece que la verdadera proporción de los pollos Perdue no contaminados difiere de aquella de la marca Tyson? Lleve a cabo una prueba de hipótesis con un nivel de significancia de .01, obteniendo un valor de *P*.

b. Si las verdaderas proporciones de los pollos no contaminados de las marcas Perdue y Tyson son .50 y .25, respectivamente, ¿qué tan probable es que la hipótesis nula de igualdad de proporciones será rechazada cuando se utiliza un nivel de significancia .01 y el tamaño de las muestras es 80?

51. Se cree que la portada y la naturaleza de la primera pregunta en encuestas por correo influyen en la proporción de respuestas. El artículo "The Impact of Cover Design and First Questions on Response Rates for a Mail Survey of Skydivers" (*Leisure Sciences*, 1991: 67–76) puso a prueba esta teoría experimentando con diferentes diseños de portadas. Una era simple, la otra utilizaba la imagen de un paracaidista. Los investigadores especularon que la proporción de respuestas *sería* más baja con la portada simple.

Portada	Números enviados	Números devueltos
Simple	207	104
Paracaidista	213	109

¿Confirman estos datos la hipótesis de los investigadores? Pruebe la hipótesis pertinente con $\alpha = .10$ calculando primero un valor P.

52. ¿Consideran los maestros que su trabajo es remunerativo y satisfactorio? El artículo "Work-Related Attitudes" (*Psychological Reports*, 1991: 443–450) reporta los resultados de una encuesta de 395 maestros de primaria y 266 maestros de preparatoria. De los maestros de primaria, 224 dijeron que estaban muy satisfechos con su trabajo, en tanto que 126 de los maestros de preparatoria estaban muy satisfechos con su trabajo. Calcule las diferencias entre la proporción de todos los maestros de primaria que están satisfechos y todos los maestros de preparatoria que están satisfechos calculando e interpretando un intervalo de confianza.

53. Olestra es un sustituto de grasa aprobado por la FDA para usarse en bocadillos. Como ha habido reportes anecdóticos de problemas gastrointestinales asociados con el consumo de olestra, se realizó un experimento de control con placebo doblemente a ciegas aleatorizado para comparar las papas fritas con olestra con las regulares con respecto a síntomas gastrointestinales ("Gastrointestinal Symptoms Following Consumption of Olestra or Regular Triglyceride Potato Chips", *J. of the Amer. Med. Assoc.*, 1998: 150–152). Entre 529 individuos en el grupo de control con las papas regulares 17.6% experimentaron un evento gastrointestinal adverso, en tanto que entre los 563 individuos en el grupo de tratamiento con papas olestra, el 15.8% experimentó dicho evento.

a. Realice una prueba de hipótesis al nivel de significancia de 5% para decidir si la proporción de incidencia de problemas gastrointestinales en aquellos que consumen papas con olestra de acuerdo con el régimen experimental difiere de la proporción de incidencia con el tratamiento de control con papas regulares.

b. Si los porcentajes verdaderos con los dos tratamientos fueron 15% y 20%, respectivamente, ¿qué tamaños de muestra ($m = n$) serían necesarios para detectar semejantes diferencias con probabilidad de .90?

54. Teen Court es un programa de diversión juvenil diseñado para eludir el tratamiento formal de los menores infractores por primera vez en el sistema de justicia de menores. El artículo "An Experimental Evaluation of Teen Courts" (*J. of Experimental Criminology*, 2008: 137–163) informó sobre un estudio en el que los delincuentes fueron asignados aleatoriamente a la Teen Court o corte de adolescentes o al Departamento de Servicios Juveniles método de procesamiento tradicional. De los 56 individuos TC, 18 posteriormente reincidieron, (¡búsquelos!) durante los 18 meses de seguimiento, mientras que 12 de los 51 individuos DJS lo hizo. ¿Los datos sugieren que la verdadera proporción de individuos TC que reincidieron durante el periodo de seguimiento especifico difieren de la proporción de individuos DJS que lo hacen? Establezca y pruebe las hipótesis pertinentes mediante la obtención de un valor P y luego con un nivel de significación de .10.

55. En investigaciones médicas, la proporción $\theta = p_1/p_2$ a menudo es de más interés que la diferencia $p_1 - p_2$ (p. ej., ¿qué tan probable es que los individuos que recibieron el tratamiento 1 se recuperen como aquellos que recibieron el tratamiento 2?) Sea $\hat{\theta} = \hat{p}_1/\hat{p}_2$. Cuando tanto m como n son grandes, el estadístico $\ln(\hat{\theta})$ tiene aproximadamente una distribución normal con valor medio aproximado $\ln(\theta)$ y desviación estándar aproximada $[(m - x)/(mx) + (n - y)/(ny)]^{1/2}$.

a. Use estos datos para obtener una fórmula para intervalo de confianza de 95% de muestra grande para calcular el $\ln(\theta)$ y luego un intervalo de confianza para θ mismo.

b. Regrese a los datos de ataque cardiaco del ejemplo 1.3 y calcule un intervalo de valores factibles de θ al nivel de confianza de 95%. ¿Qué sugiere este intervalo sobre la eficacia del tratamiento con aspirina?

56. En ocasiones algunos experimentos que implican éxitos o fallas se realizan en pares o de una manera antes/después. Suponga que antes de un discurso político importante dado por un candidato político, se seleccionan n individuos y se les preguntó si (S) o no (F) están a favor del candidato. Luego del discurso a las mismas n personas se les hizo la misma pregunta. Las respuestas pueden ser ingresadas en una tabla como sigue:

		Después	
		S	F
Antes	S	x_1	x_2
	F	x_3	x_4

donde $x_1 + x_2 + x_3 + x_4 = n$. Sean p_1, p_2, p_3 y p_4 las cuatro probabilidades de las celdas, de modo que $p_1 = P(S$ antes y S después), y así sucesivamente. Se desea probar la hipótesis de que la real proporción de simpatizantes (S) después del discurso no se ha incrementado contra la alternativa de que sí se ha incrementado.

a. Establezca las dos hipótesis de interés en función de p_1, p_2, p_3 y p_4.

b. Construya un estimador de la diferencia antes/después en probabilidades de éxito.

c. Cuando n es grande, se puede demostrar que la variable aleatoria $(X_i - X_j)/n$ tiene de manera aproximada una distribución normal con varianza dada por $[p_i + p_j - (p_i - p_j)^2]/n$. Use esto para construir un estadístico de prueba con aproximadamente una distribución estándar normal cuando H_0 es verdadera (el resultado se llama prueba de McNemar).

d. Si $x_1 = 350$, $x_2 = 150$, $x_3 = 200$ y $x_4 = 300$, ¿qué concluye?

57. Se han utilizado dos tipos diferentes de aleación, A y B, para fabricar especímenes experimentales de un pequeño eslabón sometido a tensión utilizado en una aplicación de ingeniería. Se determinó la resistencia última (kg/pulg^2) de cada espécimen y los resultados se resumen en la distribución de frecuencia adjunta.

	A	B
26 – < 30	6	4
30 – < 34	12	9
34 – < 38	15	19
38 – < 42	7	10
	$m = 40$	$m = 42$

Calcule un intervalo de confianza de 95% para la diferencia entre las proporciones verdaderas de todos los especímenes de aleaciones A y B que tienen una resistencia última de por lo menos 34 kg/pulg^2.

58. Con la fórmula tradicional, se tiene que construir un intervalo de confianza de 95% para $p_1 - p_2$ con base en tamaños de muestra iguales de las dos poblaciones. ¿Con qué valor de n $(= m)$ tendrá el intervalo resultante un ancho de cuando mucho .1 independientemente de los resultados del muestreo?

9.5 Inferencias sobre dos varianzas de población

De vez en cuando se requieren métodos para comparar dos varianzas de población (o desviaciones estándar), aunque tales problemas surgen con mucho menor frecuencia que aquellos que implican medias o proporciones. Para el caso en que las poblaciones investigadas son normales, los procedimientos están basados en una nueva familia de distribuciones de probabilidad.

La distribución *F*

La distribución de probabilidad F tiene dos parámetros, denotados por ν_1 y ν_2. El parámetro ν_1 se conoce como *número de grados de libertad del numerador* y ν_2 es el *número de grados de libertad del denominador*; en este caso ν_1 y ν_2 son enteros positivos. Una variable aleatoria que tiene una distribución F no puede asumir un valor negativo. Como la función de densidad es complicada y no será utilizada en forma explícita, se omite la fórmula. Existe una importante conexión entre una variable F y variables ji al cuadrado. Si X_1 y X_2 son variables aleatorias ji al cuadrado independientes con ν_1 y ν_2 grados de libertad, respectivamente, entonces la variable aleatoria

$$F = \frac{X_1/\nu_1}{X_2/\nu_2} \tag{9.8}$$

(la razón de las dos variables ji cuadrada divididas entre sus respectivos grados de libertad) se puede demostrar que tiene una distribución F.

La figura 9.8 ilustra la gráfica de una función de densidad F típica. Análoga a la notación $t_{\alpha,\nu}$ y $\chi^2_{\alpha,\nu}$, se utiliza F_{α,ν_1,ν_2} para el valor sobre el eje horizontal que captura α del área bajo la curva de densidad F, con ν_1 y ν_2 grados de libertad en la cola superior. La curva de densidad no es simétrica, así que parecería que tanto los valores críticos de cola

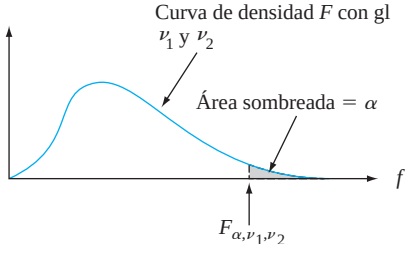


Figure 9.8 Una curva de densidad F y valor crítico

superior como los de cola inferior deben ser tabulados. Esto no es necesario, debido al hecho de que $F_{1-\alpha,\nu_1,\nu_2} = 1/F_{\alpha,\nu_2,\nu_1}$.

La tabla A.9 del apéndice da F_{α,ν_1,ν_2} con $\alpha = .10, .05, .01$, y .001 y varios valores de ν_1 (en diferentes columnas de la tabla) y ν_2 (en diferentes grupos de renglones de la tabla). Por ejemplo, $F_{.05,6,10} = 3.22$ y $F_{.05,10,6} = 4.06$. El valor crítico $F_{.95,6,10}$, que captura .95 del área a su derecha (y por tanto .05 a la izquierda) bajo la curva F con $\nu_1 = 6$ y $\nu_2 = 10$, es $F_{.95,6,10} = 1/F_{.05,10,6} = 1/4.06 = .246$

Prueba *F* para igualdad de varianzas

Un procedimiento de prueba de hipótesis que se refiere a la razón σ_1^2/σ_2^2 está basado en el siguiente resultado.

TEOREMA

Sea $X_1, \ldots, X_m$ una muestra aleatoria de una distribución normal con varianza σ_1^2, sea $Y_1, \ldots, Y_n$ otra muestra aleatoria (independiente de las X_i) de una distribución normal con varianza σ_2^2, y sean S_1^2 y S_2^2 las dos varianzas muestrales. Entonces la variable aleatoria

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \tag{9.9}$$

tiene una distribución F con $\nu_1 = m - 1$ y $\nu_2 = n - 1$.

Este teorema se obtiene al combinar (9.8) con el hecho de que cada una de las variables $(m-1)S_1^2/\sigma_1^2$ y $(n-1)S_2^2/\sigma_2^2$ tienen una distribución ji al cuadrado con $m - 1$ y $n - 1$ grados de libertad, respectivamente (véase la sección 7.4). Como F incluye una razón en lugar de una diferencia, el estadístico de prueba es la razón de varianzas muestrales. La pretensión de que $\sigma_1^2 = \sigma_2^2$ es entonces rechazada si la razón difiere en mucho de 1.

Hipótesis nula: $H_0: \sigma_1^2 = \sigma_2^2$

Valor estadístico de prueba: $f = s_1^2/s_2^2$

Hipótesis alternativa	Región de rechazo para una prueba de nivel α
$H_a: \sigma_1^2 > \sigma_2^2$	$f \geq F_{\alpha,m-1,n-1}$
$H_a: \sigma_1^2 < \sigma_2^2$	$f \leq F_{1-\alpha,m-1,n-1}$
$H_a: \sigma_1^2 \neq \sigma_2^2$	$f \geq F_{\alpha/2,m-1,n-1}$ o $f \leq F_{1-\alpha/2,m-1,n-1}$

Como los valores críticos se tabulan sólo para $\alpha = .10, .05, .01$ y .001, la prueba de dos colas se realiza sólo a los niveles .20, .10, .02 y .002. Con software estadístico se obtienen otros valores críticos F.

Ejemplo 9.14 Con base en los datos reportados en un artículo de 1979 que apareció en el *Journal of Gerontology* ("Serum Ferritin in an Elderly Population": 521–524), los autores concluyeron que la distribución de ferritina en los adultos mayores tenía un varianza más pequeña que en los adultos jóvenes (la ferritina en suero se utiliza para diagnosticar deficiencia de hierro). Para una muestra de 28 varones adultos mayores, la desviación estándar de ferritina en suero (mg/L) fue $s_1 = 52.6$; para 26 adultos jóvenes, la desviación estándar de la muestra fue $s_2 = 84.2$. ¿Confirman estos datos la conclusión tal como se aplicó a hombres?

Sean σ_1^2 y σ_2^2 las varianzas de las distribuciones de ferritina en suero para adultos mayores y adultos jóvenes, respectivamente. Se desea probar la hipótesis de interés H_0: $\sigma_1^2 = \sigma_2^2$ contra H_a: $\sigma_1^2 < \sigma_2^2$. Al nivel .01, H_0 será rechazada si $f \le F_{.99,\,27,\,25}$. Para obtener el valor crítico, se requiere $F_{.01,25,27}$. En la tabla A.9 del apéndice, $F_{.01,25,27} = 2.54$, por lo tanto $F_{.99,\,27,\,25} = 1/2.54 = .394$. El valor calculado de F es $(52.6)^2/(84.2)^2 = .390$. Como $.390 \le .394$, H_0 es rechazada al nivel .01 a favor de H_a, ya que la variabilidad parece ser más grande en adultos jóvenes que en adultos mayores. ■

Valores *P* para pruebas *F*

Recuérdese que el valor P para una prueba t de cola superior es el área bajo la curva t pertinente (aquella con grados de libertad apropiados) a la derecha de la t calculada. Del mismo modo, el valor P para una prueba F de cola superior es el área bajo la curva F con grados de libertad apropiados en el numerador y denominador a la derecha de la f calculada. La figura 9.9 ilustra esto para una prueba basada en $\nu_1 = 4$ y $\nu_2 = 6$.

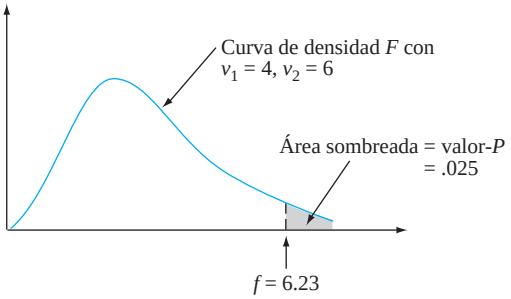


Figura 9.9 Valor *P* para una prueba *F* de cola superior

La tabulación de áreas de cola superior de una curva F es mucho más tediosa que para curvas t porque dos grados de libertad están implicados. Con cada combinación de ν_1 y ν_2, la tabla F da sólo los cuatro valores críticos que capturan las áreas .10, .05, .01 y .001. La figura 9.10 muestra lo que se puede decir sobre el valor P según dónde quede f con respecto a los cuatro valores críticos.

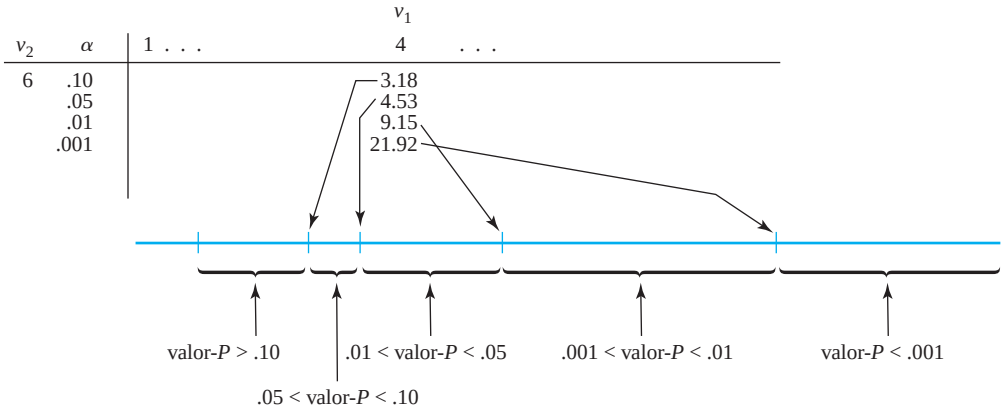


Figura 9.10 Obtención de información sobre el valor *P* en la tabla *F* para una prueba *F* de cola superior

Por ejemplo, para una prueba con $\nu_1 = 4$ y $\nu_2 = 6$,

$$f = 5.70 \Rightarrow .01 < \text{valor-}P < .05$$
$$f = 2.16 \Rightarrow \text{valor-}P > .10$$
$$f = 25.03 \Rightarrow \text{valor-}P < .001$$

Sólo si f es igual a un valor tabulado se obtiene un valor P exacto (p. ej., si $f = 4.53$, entonces el valor $P = .05$). Una vez que se sabe que $.01 <$ valor $P < .05$, H_0 sería rechazada a un nivel de significancia de .05, pero no a un nivel de .01. Cuando el valor $P < .001$, H_0 deberá ser rechazada a cualquier nivel de significancia razonable.

Todas las pruebas F discutidas en capítulos subsiguientes serán de cola superior. Si, sin embargo, una prueba F de cola inferior es apropiada, entonces, los valores críticos de cola inferior deben obtenerse como se describió antes, para que se pueda establecer un límite o límites del valor P. En el caso de una prueba de dos colas, el límite o límites de una prueba de una cola deberán ser multiplicados por 2. Por ejemplo, si $f = 5.82$ cuando $\nu_1 = 4$ y $\nu_2 = 6$, entonces puesto que 5.82 queda entre los valores críticos .05 y .01, $2(.01) <$ valor $P < 2(.05)$, es decir $.02 <$ valor $P < .10$. H_0 sería rechazada si $\alpha = .10$ pero no si $\alpha = .01$. En este caso, no se puede decir con base en la tabla qué conclusión es apropiada cuando $\alpha = .05$ (puesto que no se sabe si el valor P es más pequeño o más grande que éste). Sin embargo, el software estadístico muestra que el área a la derecha de 5.82 bajo esta curva F es .029, de modo que el valor P es .058 y que por consiguiente la hipótesis nula no debe ser rechazada al nivel .05 (.058 es el α más pequeño con el cual H_0 puede ser rechazada y el α seleccionado es más pequeño que éste). Varios programas de computadora estadísticos, desde luego, proporcionan un valor P exacto para cualquier prueba F.

Intervalo de confianza para σ_1/σ_2

El intervalo de confianza para σ_1^2/σ_2^2 se basa en el reemplazo de F en el enunciado de probabilidad

$$P(F_{1-\alpha/2,\nu_1,\nu_2} < F < F_{\alpha/2,\nu_1,\nu_2}) = 1 - \alpha$$

con la variable F (9.9) y en la manipulación de las desigualdades para aislar σ_1^2/σ_2^2. Se obtiene un intervalo para σ_1/σ_2 al tomar la raíz cuadrada de cada límite. Los detalles se dejan para un ejercicio.

EJERCICIOS Sección 9.5 (59–66)

59. Obtenga o calcule las siguientes cantidades:
- **a.** $F_{.05,5,8}$ **b.** $F_{.05,8,5}$ **c.** $F_{.95,5,8}$ **d.** $F_{.95,8,5}$
- **e.** El 99avo percentil de la distribución F con $\nu_1 = 10$, $\nu_2 = 12$
- **f.** El 1er percentil de la distribución F con $\nu_1 = 10$, $\nu_2 = 12$
- **g.** $P(F \le 6.16)$ para $\nu_1 = 6$, $\nu_2 = 4$
- **h.** $P(.177 \le F \le 4.74)$ para $\nu_1 = 10$, $\nu_2 = 5$

60. Dé tanta información como pueda sobre el valor P de la prueba F en cada una de las siguientes situaciones:
- **a.** $\nu_1 = 5$, $\nu_2 = 10$, prueba de cola superior, $f = 4.75$
- **b.** $\nu_1 = 5$, $\nu_2 = 10$, prueba de cola superior, $f = 2.00$
- **c.** $\nu_1 = 5$, $\nu_2 = 10$, prueba de dos colas, $f = 5.64$
- **d.** $\nu_1 = 5$, $\nu_2 = 10$, prueba de cola inferior, $f = .200$
- **e.** $\nu_1 = 35$, $\nu_2 = 20$, prueba de cola superior, $f = 3.24$

61. Regrese a los datos sobre ángulo de inclinación máximo dados en el ejercicio 28 de este capítulo. Realice una prueba a nivel de significación de .10 para ver si las desviaciones estándar de población de los dos grupos de edad son diferentes (las curvas de probabilidad normal confirman la suposición de normalidad necesaria).

62. Remítase al ejemplo 9.7. ¿Sugieren los datos que la desviación estándar de la distribución de resistencia de especímenes sometidos a un proceso de fusión es más pequeña que la de especímenes no sometidos a un proceso de fusión? Realice una prueba de significancia de .01 obteniendo tanta información como pueda sobre el valor P.

63. El toxafen es un insecticida que ha sido identificado como contaminante en el ecosistema de los Grandes Lagos. Para investigar el efecto de la exposición al toxafen en animales, a grupos

de ratas se les administró toxafen en su dieta. El artículo "Reproduction Study of Toxaphene in the Rat" (*J. of Environ. Sci. Health*, 1988: 101–126) reporta aumentos de peso (en gramos) de ratas a las que se les administró una dosis baja (4 ppm) y de ratas de control cuya dieta no incluía el insecticida. La desviación estándar de muestra de 23 ratas hembra de control fue de 32 g y de 20 ratas hembra sometidas a dosis bajas fue de 54 g. ¿Sugieren estos datos que existe más variabilidad en los incrementos de peso a dosis bajas que en los incrementos de peso en las ratas de control? Suponiendo normalidad, realice una prueba de hipótesis a nivel de significación de .05.

64. En un estudio de deficiencia de cobre en ganado, se determinaron los valores de cobre (μg Cu/100 ml de sangre) tanto para ganado que pace en un área donde se sabe que existen anomalías bien definidas provocadas por molibdeno (valores de contenido del metal que exceden el rango normal de variación regional) y para ganado que pace en área no anómala. ("An Investigation into Copper Deficiency in Cattle in the Southern Pennines", *J. Agricultural Soc. Cambridge*, 1972: 157–163), con el resultado $s_1 = 21.5$ ($m = 48$) en la condición anómala y $s_2 = 19.45$ ($n = 45$) para la condición no anómala. Pruebe en cuanto a igualdad contra desigualdad de varianzas de población a un nivel de significación de .10 utilizando el método del valor *P*.

65. El artículo "Enhancement of Compressive Properties of Failed Concrete Cylinders with Polymer Impregnation" (*J. of Testing and Evaluation*, 1977: 333–337) reporta los siguientes datos sobre módulo de compresión impregnado (lb/pulg^2 $\times$ 10^6) cuando se utilizaron dos polímeros diferentes para reparar grietas en concreto que falló.

Epoxi	1.75	2.12	2.05	1.97
MMA prepolímero	1.77	1.59	1.70	1.69

Obtenga un intervalo de confianza de 90% para la proporción de variaciones aplicando primero el método sugerido en el libro para obtener la fórmula general para un intervalo de confianza.

66. Reconsidere los datos del ejemplo 9.6 y calcule un límite de confianza superior de 95% para la razón de la desviación estándar de la distribución de porosidad en triacetato respecto a la de la distribución de porosidad en algodón.

EJERCICIOS SUPLEMENTARIOS (67–95)

67. Los datos adjuntos sobre resistencia a la compresión (lb) de cajas de 12 $\times$ 10 $\times$ 8 pulg aparecieron en el artículo "Compression of Single-Wall Corrugated Shipping Containers Using Fixed and Floating Test Platens" (*J. Testing and Evaluation*, 1992: 318–320). Los autores manifestaron que "la diferencia entre la resistencia a la compresión utilizando un método de platinas fijas y flotantes es pequeña comparada con la variación normal de la resistencia a la compresión entre cajas idénticas". ¿Está de acuerdo? ¿Su análisis se basa en cualquier suposición?

Método	Tamaño muestral	Media muestral	Desv. est. muestral
Fijo	10	807	27
Flotante	10	757	41

68. Los autores del artículo "Dynamics of Canopy Structure and Light Interception in *Pinus elliotti*, North Florida" (*Ecological Monographs*, 1991: 33–51) idearon un experimento para determinar el efecto de un fertilizante en un área cubierta de hojas. Se dispuso de varios solares para el estudio y se seleccionó al azar la mitad para fertilizarlos. Para asegurarse de que los solares que iban a recibir el fertilizante y los de control fueran iguales antes de comenzar el experimento se registró la densidad de árboles (el número de árboles por hectárea) en ocho solares que iban a ser fertilizados y en ocho solares de control y se obtuvieron los resultados siguientes generados por Minitab.

Solares fertilizados	1024	1216	1312	1280
	1216	1312	992	1120
Solares de control	1104	1072	1088	1328
	1376	1280	1120	1200

```
Dos muestras T para fertilizante vs control

               N    Media      DE   EE Media
fertilizantes  8     1184     126         44
control        8     1196     118         42

95% IC para fertilizante mu − control mu:
(−144, 120)
```

a. Construya una gráfica de caja comparativa y comente sobre cualquier característica interesante.

b. ¿Concluiría que existe una diferencia significativa en la densidad de árboles media en los solares fertilizados y de control? Use $\alpha = .05$.

c. Interprete el intervalo de confianza dado.

69. ¿Se ve afectada la proporción de respuestas a cuestionarios si se incluye un incentivo por responder junto con el cuestionario? en un experimento, de 110 cuestionarios sin incentivo 75 fueron regresados, en tanto que 98 cuestionarios que incluían la oportunidad de ganar un premio de lotería dieron por resultado 66 respuestas ("Charities, No; Lotteries, No; Cash, Yes", *Public Opinion Quarterly*, 1996: 542–562). ¿Sugieren estos datos que la inclusión de un incentivo incrementa la probabilidad de una respuesta? Formule y pruebe las hipótesis pertinentes a un nivel de significación de .10 utilizando el método del valor *P*.

70. Los datos adjuntos se obtuvieron en un estudio para evaluar el potencial de licuefacción en una planta de energía nuclear propuesta ("Cyclic Strengths Compared for Two Sampling Techniques", *J. of the Geotechnical Division, Amer. Soc. Civil Engrs. Proceedings*, 1981: (563–576). Antes de probar la resistencia cíclica, se recopilaron muestras de suelo mediante un método de jarro y un método de bloque y se obtuvieron los siguientes valores observados de densidad en seco (lb/pie^3):

Muestreo con jarro	101.1	111.1	107.6	98.1
	99.5	98.7	103.3	108.9
	109.1	104.1	110.0	98.4
	105.1	104.5	105.7	103.3
	100.3	102.6	101.7	105.4
	99.6	103.3	102.1	104.3
Muestreo con bloque	107.1	105.0	98.0	97.9
	103.3	104.6	100.1	98.2
	97.9	103.2	96.9	

Calcule e interprete un intervalo de confianza de 95% para la diferencia entre densidades en seco promedio verdaderas para los dos métodos de muestreo.

71. El artículo "Quantitative MRI and Electrophysiology of Preoperative Carpal Tunnel Syndrome in a Female Population" (*Ergonomics*, 1997: 642–649) reportó que $(-473.3, 1691.9)$ era un intervalo de confianza de 95% con muestra grande para la diferencia entre el promedio real del volumen del músculo ténar (mm^3) en personas que padecen el síndrome de túnel carpiano y el volumen promedio verdadero en personas que no padecen el síndrome. Calcule un intervalo de confianza de 90% para esta diferencia.

72. Los siguientes datos de resistencia a la flexión (lb-pulg/pulg) de juntas se tomaron del artículo "Bending Strength of Corner Joints Constructed with Injection Molded Splines" (*Forest Products J.*, abril de 1997: 89–92).

Tipo	Tamaño de muestra	Media muestral	DE muestral
Sin recubrimiento lateral	10	80.95	9.59
Con recubrimiento lateral	10	63.23	5.96

a. Calcule un límite de confianza inferior de 95% para la resistencia promedio verdadera de juntas con recubrimiento lateral.
b. Calcule un límite de predicción inferior de 95% para la resistencia de una sola junta con recubrimiento lateral.
c. Calcule un intervalo que, con confianza de 95%, incluya los valores de resistencia de por lo menos 95% de la población de todas las juntas con recubrimientos laterales.
d. Calcule un intervalo de confianza de 95% para la diferencia entre las resistencias promedio verdaderas de los dos tipos de juntas.

73. El artículo "Urban Battery Litter", citado en el ejemplo 8.14 dio el siguiente resumen de datos sobre la masa de zinc (g) para dos marcas diferentes de baterías de tamaño D:

Marca	Tamaño muestral	Media muestral	Desv. est. muestral
Duracell	15	138.52	7.76
Energizer	20	149.07	1.52

Suponiendo que ambas distribuciones de masa de zinc son por lo menos aproximadamente normales, lleve a cabo una prueba al nivel de significación .05 utilizando el método de valor *P* para decidir si las masas reales de zinc promedio son diferentes para las dos marcas de baterías.

74. El descarrilamiento de un tren de carga provocado por la catastrófica falla de la chumacera de la armadura de un motor de tracción motivó un estudio reportado en el artículo "Locomotive Traction Motor Armature Bearing Life Study" (*Lubrication Engr.*, agosto de 1997: 12–19). Se seleccionó una muestra de 17 motores de tracción con alto kilometraje y se determinó la penetración de cono (mm/10) tanto en la chumacera de piñón como en la chumacera de la armadura del conmutador y se obtuvieron los siguientes datos:

	Motor					
	1	**2**	**3**	**4**	**5**	**6**
Conmutador	211	273	305	258	270	209
Piñón	226	278	259	244	273	236

	Motor					
	7	**8**	**9**	**10**	**11**	**12**
Conmutador	223	288	296	233	262	291
Piñón	290	287	315	242	288	242

	Motor				
	13	**14**	**15**	**16**	**17**
Conmutador	278	275	210	272	264
Piñón	278	208	281	274	268

Calcule la diferencia media de población entre la penetración en la chumacera de la armadura del conmutador y la penetración en la chumacera de piñón y hágalo de manera que permita obtener información sobre confiabilidad y precisión de la estimación. [*Nota*: una curva de probabilidad normal valida la suposición de normalidad necesaria.] ¿Diría que la diferencia media de la población ha sido estimada con precisión? ¿Difiere la penetración media de población para los dos tipos de chumacera? Explique.

75. La "cabezabilidad" es la capacidad de una pieza cilíndrica de permitir que se le dé la forma de la cabeza de un perno, tornillo u otra parte formada en frío sin rotura. El artículo "New Methods for Assessing Cold Heading Quality" *(Wire J. Intl.*, octubre de 1996: 66–72) describe el resultado de una prueba de impacto de cabezabilidad aplicada a 30 especímenes de acero muerto al aluminio y a 30 especímenes de acero muerto al silicio. El número de clasificación de cabezabilidad medio de la muestra para los especímenes de acero fue de 6.43 y la media de la muestra para los especímenes de aluminio fue de 7.09. Suponga que las desviaciones estándar de la muestra fueran de 1.08 y 1.19, respectivamente. ¿Está de acuerdo con los autores del artículo en que la diferencia en las clasificaciones de cabezabilidad es significativa al nivel de 5% (suponiendo que las dos distribuciones de cabezabilidad son normales?

76. El artículo "Fatigue Testing of Condoms", citado en el ejercicio 7.32 informó que para una muestra de 20 condones de látex natural de un cierto tipo, la media muestral y la desviación estándar de la muestra del número de ciclos de rompimiento fueron 4358 y 2218, respectivamente, mientras que una muestra de 20 condones de poliisopreno dio una media y una desviación estándar de la muestra de 5805 y 3990, respectivamente. ¿Hay una fuerte evidencia para concluir que el verdadero número promedio de ciclos de ruptura para el preservativo de poliisopreno supera al de los condones de látex natural en más de 1000 ciclos? Lleve a cabo una prueba con un nivel de signi-

ficación de .01. [*Nota*: el artículo citado informó valores *P* de las pruebas *t* para comparar las medias de los diversos tipos considerados.]

77. Se requiere información sobre la postura de la mano y las fuerzas generadas por los dedos durante la manipulación de varios objetos cotidianos para diseñar prótesis de alta tecnología para la mano. El artículo "Grip Posture and Forces During Holding Cylindrical Objects with Circular Grips" (*Ergonomics*, 1996: 1163–1176) reportó que para una muestra de 11 mujeres, la fuerza media de opresión con cuatro dedos (N) fue de 98.1 y la desviación estándar de la muestra fue de 14.2. Para una muestra de 15 hombres, la media de la muestra y la desviación estándar de la muestra fueron de 129.2 y 39.1, respectivamente.

a. Una prueba realizada para ver si las fuerzas promedio verdaderas para los dos géneros eran diferentes dio por resultado $t = 2.51$ y valor $P = .019$. ¿Da el procedimiento de prueba apropiado descrito en este capítulo este valor de *t* y el valor *P* establecido?

b. ¿Existe evidencia sustancial para concluir que la fuerza promedio verdadera para hombres excede la de la fuerza de mujeres por más de 25 N? Formule y pruebe las hipótesis pertinentes.

78. El artículo "Pine Needles as Sensors of Atmospheric Pollution" (*Environ. Monitoring*, 1982: 273–286) reportó sobre el uso de análisis de actividad de neutrones para determinar concentración de contaminantes en hojas de pino. De acuerdo con los autores del artículo, "estas observaciones indican fuertemente que para aquellos elementos bien determinados mediante procedimientos analíticos, la distribución de concentración es lognormal. Por consiguiente, en pruebas de significancia se utilizarán logaritmos de concentraciones". Los datos dados se refieren a concentración de bromo en hojas de pino tomadas del sitio cercano a una planta de vapor alimentada con petróleo y de un sitio relativamente limpio. Los valores dados a continuación son medias y desviaciones estándar de las observaciones logarítmicas transformadas.

Sitio	Tamaño de muestra	Concentración log media	Desv. est. de concentración/log
Planta de vapor	8	18.0	4.9
Limpia	9	11.0	4.6

Sea μ_1^* la concentración *log* promedio verdadera en el primer sitio y defínase μ_2^* análogamente para el segundo sitio.

a. Use la prueba *t* agrupada (basada en la suposición de normalidad *y* desviaciones estándar iguales) para decidir un nivel de significancia de .05 si las dos medias de distribución de concentración son iguales.

b. Si σ_1^* y σ_2^* (las desviaciones estándar de las dos distribuciones log de concentración), no son iguales, ¿serían μ_1 y μ_2 (las medias de las distribuciones de concentración), las mismas si $\mu_1^* = \mu_2^*$? Explique su razonamiento.

79. El artículo "The Accuracy of Stated Energy Contents of Reduced-Energy, Commercially Prepared Foods" *(J. of the Amer. Dietetic Assoc.*, 2010: 116–123) presentó los datos que acompañan a la energía bruta del proveedor indicado y el valor medido (ambos en kcal) para 10 comidas de diferentes supermercados de conveniencia):

Comida:	1	2	3	4	5	6	7	8	9	10
Declarada:	180	220	190	230	200	370	250	240	80	180
Medida:	212	319	231	306	211	431	288	265	145	228

Lleve a cabo una prueba de hipótesis basada en un valor *P* para decidir si la verdadera diferencia promedio % que se declaró difiere de cero. [*Nota*: el artículo declaró: "Aunque los métodos estadísticos formales no se aplican a muestras de conveniencia, las pruebas estándar de estadística se emplearon para resumir los datos con fines de exploración y sugerir direcciones para futuros estudios."]

80. El arsénico es un carcinógeno conocido y un veneno. Los procedimientos estándar de laboratorio para medir la concentración de arsénico (μg/L) en el agua son caros. Considere el resumen que acompaña los datos de entrada y salida de Minitab para comparar un método de laboratorio con un método de campo relativamente nuevo, rápido y barato (tomado del artículo "Evaluation of a New Field Measurement Method for Arsenic in Drinking Water Samples", *J. of Envir. Engr.*, 2008: 382–388).

Prueba *t* e IC para dos muestras

```
Muestra   N    Media   Desv Est   Desv. est.
                                    muestral
1         3    19.70       1.10        0.64
2         3    10.90       0.60        0.35
Diferencia estimada: 8.800
95% IC para la diferencia: (6.498, 11.102)
Prueba T de la diferencia = 0 (vs no =):
Valor T = 12.16 Valor P = 0.001 GL = 3
```

¿Qué conclusión saca usted sobre los dos métodos y por qué? Interprete el intervalo de confianza dado. [*Nota*: uno de los autores del artículo indica en una comunicación privada que no estaban seguros de por qué los dos métodos no concordaban.]

81. Los datos que acompañan al tiempo de respuesta aparecieron en el artículo "The Extinguishment of Fires Using Low-Flow Water Hose Streams—Part II" (*Fire Technology*, 1991; 291–320).

Buena visibilidad
.43 1.17 .37 .47 .68 .58 .50 2.75

Mala visibilidad
1.47 .80 1.58 1.53 4.33 4.23 3.25 3.22

Los autores analizaron los datos con la prueba *t* agrupada. ¿El uso de esta prueba parece justificado? [*Sugerencia*: compruebe la normalidad. Los percentiles *z* para $n = 8$ son -1.53, $-.89$, $-.49$, $-.15$, .15, .49, .89 y 1.53.]

82. Comúnmente se utiliza cemento acrílico para hueso en la artroplastia total de articulaciones como un material que permite la transferencia suave de cargas de una prótesis metálica a una estructura ósea. El artículo "Validation of the Small-Punch Test as a Technique for Characterizing the Mechanical Properties of Acrylic Bone Cement" (*J. of Engr. in Med.*, 2006: 11–21) dio los siguientes datos sobre fuerza de ruptura (N):

Temp	Medio	*n*	$\bar{x}$	*s*
22°	Seco	6	170.60	39.08
37°	Seco	6	325.73	34.97
22°	Húmedo	6	366.36	34.82
37°	Húmedo	6	306.09	41.97

Suponga que todas las distribuciones de población son normales.

a. Estime la fuerza de ruptura promedio verdadera en un medio seco a 37° y hágalo de una forma que dé información sobre precisión y confiabilidad. Luego interprete su estimación.

b. Estime la diferencia entre la fuerza a la ruptura promedio verdadera en un medio seco a 37° y la fuerza promedio verdadera a la misma temperatura en un medio húmedo y hágalo de modo que obtenga información sobre precisión y confiabilidad. Luego interprete su estimación.

c. ¿Existe una fuerte evidencia para concluir que la fuerza promedio verdadera en un medio seco a la temperatura más alta excede a la de la temperatura más baja por más de 100 N?

83. En un experimento para comparar resistencias de apoyo de clavijas insertadas en dos tipos diferentes de soportes de montaje, una muestra de 14 observaciones de límite de esfuerzo de soportes de montaje de roble rojo dieron por resultado una media y desviación estándar muestral de 8.48 MPa y .79 MPa, respectivamente, en tanto que una muestra de 12 observaciones cuando se utilizaron soportes de montaje de abeto Douglas dieron una media de 9.36 y una desviación estándar de 1.52 ("Bearing Strength of White Oak Pegs in Red Oak and Douglas Fir Timbers", *J. of Testing and Evaluation*, 1998, 109–114). Considere probar si los límites de esfuerzo promedio verdaderos son o no idénticos para los dos tipos de soporte de montaje. Compare los grados de libertad y valores *P* para las pruebas *t* agrupadas y no agrupadas.

84. ¿Cómo se compara la absorción de energía con el consumo de energía? Un aspecto de este tema se consideró en el artículo "Measurement of Total Energy Expenditure by the Doubly Labelled Water Method in Professional Soccer Players" (*J. of Sports Sciences*, 2002: 391–397), el cual contiene los datos adjuntos (MJ/día).

	Jugador						
	1	**2**	**3**	**4**	**5**	**6**	**7**
Consumo	14.4	12.1	14.3	14.2	15.2	15.5	17.8
Absorción	14.6	9.2	11.8	11.6	12.7	15.0	16.3

Pruebe si existe una diferencia significativa entre absorción y consumo. ¿Depende la conclusión de si se utiliza un nivel de significación de .05, .01 o .001?

85. Un experimentador desea obtener un intervalo de confianza para la diferencia entre resistencia a la ruptura promedio verdadera para cables fabricados por la compañía I y la compañía II. Suponga que la resistencia a la ruptura está normalmente distribuida para ambos tipos de cable con $\sigma_1 = 30$ lb/pulg^2 y $\sigma_2 = 20$ lb/pulg^2.

a. Si los costos dictan que el tamaño de muestra para el cable de tipo I deberá ser tres veces el tamaño de muestra para el cable de tipo II, ¿cuántas observaciones se requieren si el intervalo de confianza de 99% no debe ser más ancho que 20 lb/pulg^2?

b. Suponga que se tienen que hacer un total de 400 observaciones. ¿Cuántas de ellas deberán ser hechas en muestras de cable de tipo I si el ancho del intervalo resultante tiene que ser mínimo?

86. En el artículo "Development Rates and a Temperature-Dependent Model of Pales Weevil" (*Environ. Entomology*, 1987: 956–962) se describe un experimento para determinar los efectos de la temperatura en la sobrevivencia de huevos de insectos. A 11°C, 73 de 91 huevos sobrevivieron hasta la siguiente etapa de desarrollo. A 30°C, 102 de 110 huevos sobrevivieron. ¿Sugieren los resultados de este experimento que la tasa de sobrevivencia (proporción de sobrevivencia) difiere para las dos temperaturas? Calcule el valor *P* y utilícelo para probar las hipótesis apropiadas.

87. Los meseros en restaurantes han empleado varias estrategias para incrementar las propinas. Un artículo en el ejemplar del 5 de septiembre de 2005 del *New Yorker* reportó que: "En un estudio una mesera recibió 50% más en propinas cuando se presentó con su nombre que cuando no lo hizo". Considere los siguientes datos (ficticios) sobre la cantidad de propina como un porcentaje de la cuenta.

Con presentación: $m = 50$ $\bar{x} = 22.63$ $s_1 = 7.82$

Sin presentación: $n = 50$ $\bar{y} = 14.15$ $s_2 = 6.10$

¿Sugieren estos datos que una presentación incrementa las propinas en promedio en más de 50%? Formule y pruebe las hipótesis pertinentes. [*Sugerencia*: considere el parámetro $\theta = \mu_1 - 1.5\mu_2$.]

88. El artículo "Quantitative Assessment of Glenohumeral Translation in Baseball Players" (*The Amer. J. of Sports Med.*, 2004: 1711–1715) consideró varios aspectos de movimiento del hombro para una muestra de pitchers (lanzadores) y otra muestra de jugadores de campo (*glenohumeral* se refiere a la articulación entre el húmero (bola) y el glenoide (cuenca). Los autores amablemente proporcionaron los siguientes datos sobre traslación anteroposterior (mm), una medida de la extensión del movimiento anterior y posterior, tanto del brazo dominante como del brazo no dominante.

	Pos Dom Tr	Pos ND Tr	Pit Dom Tr	Pit ND Tr
1	30.31	32.54	27.63	24.33
2	44.86	40.95	30.57	26.36
3	22.09	23.48	32.62	30.62
4	31.26	31.11	39.79	33.74
5	28.07	28.75	28.50	29.84
6	31.93	29.32	26.70	26.71
7	34.68	34.79	30.34	26.45
8	29.10	28.87	28.69	21.49
9	25.51	27.59	31.19	20.82
10	22.49	21.01	36.00	21.75
11	28.74	30.31	31.58	28.32
12	27.89	27.92	32.55	27.22
13	28.48	27.85	29.56	28.86
14	25.60	24.95	28.64	28.58
15	20.21	21.59	28.58	27.15
16	33.77	32.48	31.99	29.46
17	32.59	32.48	27.16	21.26
18	32.60	31.61		
19	29.30	27.46		
media	29.4463	29.2137	30.7112	26.6447
Desv. Est.	5.4655	4.7013	3.3310	3.6679

a. Estime la diferencia promedio verdadera de traslación entre los brazos dominante y no dominante de lanzadores en una forma que aporte información sobre confiabilidad y precisión e interprete la estimación resultante.

b. Repita (a) para jugadores de campo.

c. Los autores afirmaron que los "lanzadores mostraron una mayor diferencia en la traslación anteroposterior de lado a lado de sus hombros en comparación con los jugadores de campo". ¿Está de acuerdo? Explique.

89. Suponga que se tiene que realizar una prueba al nivel .05 de H_0: $\mu_1 - \mu_2 = 0$ contra H_a: $\mu_1 - \mu_2 > 0$, suponiendo $\sigma_1 = \sigma_2 = 10$ y normalidad para ambas distribuciones, utilizando tamaños de muestra iguales ($m = n$). Evalúe la probabilidad de un tipo de error II cuando $\mu_1 - \mu_2 = 1$ y $n = 25, 100, 2500$ y 10,000. ¿Puede pensar en problemas reales en los cuales la diferencia $\mu_1 - \mu_2 = 1$ tiene poca significación práctica? ¿Serían deseables tamaños de muestra de $n = 10{,}000$ en tales problemas?

90. Los siguientes datos se refieren a la cuenta de bacterias transportadas por el aire (número de colonias/pie^3) tanto para $m = 8$ cuartos de hospital alfombrados como para $n = 8$ cuartos no alfombrados ("Microbial Air Sampling in a Carpeted Hospital", *J. of Environmental Health*, 1968: 405). ¿Parece haber una diferencia en el conteo de bacterias promedio verdadero entre cuartos alfombrados y no alfombrados?

Alfombrado	11.8	8.2	7.1	13.0	10.8	10.1	14.6	14.0
No alfombrado	12.1	8.3	3.8	7.2	12.0	11.1	10.1	13.7

Suponga que posteriormente se dio cuenta que los cuartos alfombrados estaban en un hospital de veteranos, en tanto que los no alfombrados estaban en un hospital infantil. ¿Sería capaz de evaluar el efecto del alfombrado? Comente.

91. Investigadores enviaron 5000 currículos en respuesta a anuncios de trabajo que aparecieron en el *Boston Globe* y el *Chicago Tribune*. Los currículos eran idénticos excepto que 2500 de ellos tenían apellidos "que sonaban a apellidos de persona blanca", tales como Brett y Emily, en tanto que los otros 2500 tenían nombres "que sonaban a persona negra" tales como Tamika y Rasheed. Los currículos del primer tipo produjeron 250 respuestas y los del segundo tipo sólo 167 respuestas (estos números son muy consistentes con la información que apareció en un reporte del 15 de enero de 2003 de la Associated Press). ¿Sugieren fuertemente estos datos que un currículo con un apellido de "negro" es menos probable que dé por resultado una respuesta que un currículo con un apellido de "blanco"?

92. La prueba de McNemar, desarrollada en el ejercicio 54, también puede ser utilizada cuando los individuos son reunidos en pares para obtener n pares y luego un miembro de cada par recibe el tratamiento 1 y el otro el tratamiento 2. Luego X_1 es el número de pares en los que ambos tratamientos fueron exitosos y asimismo para X_2, X_3, X_4. El estadístico de prueba para comprobar la eficacia de los dos tratamientos está dado por $(X_2 - X_3)/\sqrt{(X_2 + X_3)}$, el cual tiene aproximadamente una distribución normal estándar cuando H_0 es verdadera. Úselo para probar si la ergotamina es efectiva en el tratamiento de dolores de cabeza de migraña.

		Ergotamina	
		S	***F***
Placebo	***S***	44	34
	F	46	30

Los datos son ficticios, pero la conclusión concuerda con la del artículo "Controlled Clinical Trial of Ergotamine Tartrate" (*British Med. J.*, 1970: 325–327).

93. El artículo "Evaluating Variability in Filling Operations" (*Food Tech.*, 1984: 51–55) describe dos operaciones de llenado diferentes utilizadas en una planta empacadora de carne molida. Ambas operaciones de llenado se ajustaron para llenar paquetes con 1400 g de carne molida. En una muestra aleatoria de tamaño 30 tomada de cada operación de llenado, las medias y desviaciones estándar resultantes fueron 1402.24 g y 10.97 g para la operación 1 y 1419.63 g y 9.96 g para la operación 2.

a. Con un nivel de significancia de .05, ¿existe suficiente evidencia que indique que el peso medio verdadero de los paquetes difiere para las dos operaciones?

b. ¿Sugieren los datos de la operación 1 que el peso medio verdadero de los paquetes producidos por la operación 1 es más grande por más de 1400 g? Use un nivel de significancia de .05.

94. Sean $X_1, \ldots, X_m$ una muestra aleatoria de una distribución de Poisson con parámetro μ_1 y sea $Y_1, \ldots, Y_n$ una muestra aleatoria de otra distribución de Poisson con parámetro μ_2. Se desea probar H_0: $\mu_1 - \mu_2 = 0$ contra una de las tres alternativas estándar. Cuando m y n son grandes se puede utilizar la prueba z con muestra grande de la sección 9.1. Sin embargo, el hecho de que $V(\overline{X}) = \mu/n$ sugiere que se debe utilizar un denominador diferente al estandarizar $\overline{X} - \overline{Y}$. Desarrolle un procedimiento de prueba con muestra grande apropiado para este problema y luego aplíquelo a los siguientes datos para probar si las densidades de plantas de una especie particular son iguales en dos regiones diferentes (donde cada observación es el número de plantas encontradas en un rectángulo de muestreo localizado al azar con área de 1 m^2, así que en la región 1 hubo 40 rectángulos en los que se observó una planta, etc.):

	Frecuencia								
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	
Región 1	28	40	28	17	8	2	1	1	$m = 125$
Región 2	14	25	30	18	49	2	1	1	$n = 140$

95. Remitiéndose al ejercicio 94, desarrolle una fórmula para el intervalo de confianza con muestra grande para $\mu_1 - \mu_2$. Calcule el intervalo para los datos dados allí utilizando un nivel de confianza de 95%.

Bibliografía

Véase la bibliografía al final del capítulo 7.

10 Análisis de la varianza

INTRODUCCIÓN

Al estudiar los métodos de análisis de datos cuantitativos, primero se trataron problemas que implican una sola muestra de números y luego se abordó el análisis comparativo de dos muestras diferentes. En problemas de una muestra, los datos se componían de observaciones sobre o respuestas de individuos u objetos experimentales seleccionados de una sola población. En problemas de dos muestras, las dos muestras se tomaron de dos poblaciones diferentes y los parámetros de interés fueron las medias de población o bien se aplicaron dos tratamientos distintos a unidades experimentales (individuos u objetos) seleccionados de una sola población; en el último caso, los parámetros de interés fueron las medias de tratamiento verdaderas.

El **análisis de la varianza**, o más brevemente, **ANOVA**, se refiere en general a un conjunto de situaciones experimentales y procedimientos estadísticos para el análisis de respuestas cuantitativas de unidades experimentales. El problema ANOVA más simple se conoce indistintamente como **unifactorial, de clasificación única** o **ANOVA unidireccional** e implica el análisis de datos muestreados de más dos poblaciones (distribuciones) numéricas o de datos de experimentos en los cuales se utilizaron más de dos tratamientos. La característica que diferencia los tratamientos o poblaciones una de otra se llama **factor** en estudio y los distintos tratamientos o poblaciones se conocen como **niveles** del factor. Ejemplos de tales situaciones incluyen los siguientes:

1. Un experimento para estudiar los efectos de cinco marcas diferentes de gasolina con respecto a la eficiencia de operación de un motor automotriz (mpg)
2. Un experimento para estudiar los efectos de la presencia de cuatro soluciones azucaradas diferentes (glucosa, sucrosa, fructosa y una mezcla de las tres) en cuanto a crecimiento de bacterias

3. Un experimento para investigar si la concentración de madera dura en la pulpa (%) afecta la resistencia a la tensión de bolsas hechas de la pulpa
4. Un experimento para decidir si la densidad de color de un espécimen de tela depende de la cantidad de tinte utilizado

En (1) el factor de interés es la marca de la gasolina y existen cinco niveles diferentes del factor. En (2) el factor es el azúcar con cuatro niveles (o cinco, si se utiliza una solución de control que no contenga azúcar). Tanto en (1) como en (2), el factor es de naturaleza cualitativa y los niveles corresponden a posibles categorías del factor. En (3) y (4), los factores son concentración de madera dura y cantidad de tinte, respectivamente; estos dos factores son de naturaleza cuantitativa, por lo que los niveles identifican diferentes ajustes del factor. Cuando el factor de interés es cuantitativo, también se pueden utilizar técnicas estadísticas de análisis de regresión (discutido en los capítulos 12 y 13) para analizar los datos.

Este capítulo se enfoca en el ANOVA unifactorial. La sección 10.1 presenta la prueba F para demostrar la hipótesis nula de que las medias de población o tratamiento son idénticas. La sección 10.2 considera un análisis adicional de los datos cuando H_0 ha sido rechazada. La sección 10.3 se ocupa de algunos otros aspectos del ANOVA unifactorial. El capítulo 11 introduce experimentos ANOVA que implican más de un factor.

10.1 ANOVA unifactorial

El ANOVA unifactorial se enfoca en la comparación de más de dos medias de población o tratamiento. Sean

I = el número de poblaciones o tratamientos que se están comparando

μ_1 = la media de población 1 o la respuesta promedio verdadera cuando se aplica el tratamiento 1

$\vdots$

μ_I = la media de población I o la respuesta promedio verdadera cuando se aplica el tratamiento I

Las hipótesis pertinentes son

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_I$$

contra

H_a: por lo menos dos de las μ_i son diferentes

Si $I = 4$, H_0 es verdadera sólo si las cuatro μ_i son idénticas. H_a sería verdadera, por ejemplo, si $\mu_1 = \mu_2 \neq \mu_3 = \mu_4$, si $\mu_1 = \mu_3 = \mu_4 \neq \mu_2$, o si las cuatro μ_i difieren una de otra.

Una prueba de estas hipótesis requiere que se tenga disponible una muestra aleatoria de cada población o tratamiento.

Ejemplo 10.1 El artículo "Compression of Single-Wall Corrugated Shipping Containers Using Fixed and Floating Test Platens" (*J. Testing and Evaluation*, 1992: 318-320) describe un experimento en el cual se compararon varios tipos diferentes de cajas con respecto a resistencia a la

compresión (lb). La tabla 10.1 presenta los resultados de un experimento ANOVA unifactorial que implica $I = 4$ tipos de cajas (las medias y desviaciones estándar muestrales concuerdan con los valores dados en el artículo).

Tabla 10.1 Datos y cantidades resumidas para el ejemplo 10.1

Tipo de caja	Resistencia a la compresión (lb)	Media muestral	DE muestral
1	655.5 788.3 734.3 721.4 679.1 699.4	713.00	46.55
2	789.2 772.5 786.9 686.1 732.1 774.8	756.93	40.34
3	737.1 639.0 696.3 671.7 717.2 727.1	698.07	37.20
4	535.1 628.7 542.4 559.0 586.9 520.0	562.02	39.87
	Media grande =	682.50	

Con μ_i denotando la resistencia a la compresión promedio verdadera de las cajas de tipo i ($i = 1, 2, 3, 4$), la hipótesis nula es H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$. La figura 10.1(a) muestra una gráfica de caja comparativa para las cuatro muestras. Existe una cantidad sustancial de traslape entre las observaciones de los primeros tres tipos de cajas, pero las resistencias a la compresión del cuarto tipo parecen considerablemente más pequeñas que para los demás tipos. Esto sugiere que H_0 no es verdadera. La gráfica de caja comparativa que aparece en la figura 10.1(b) está basada en agregar 120 a cada observación en la cuarta muestra (y así se obtiene una media de 682.02 y la misma desviación estándar) y las demás observaciones no cambian. Ya no es obvio si H_0 es verdadera o falsa. En situaciones como ésta, se requiere un procedimiento de prueba formal.

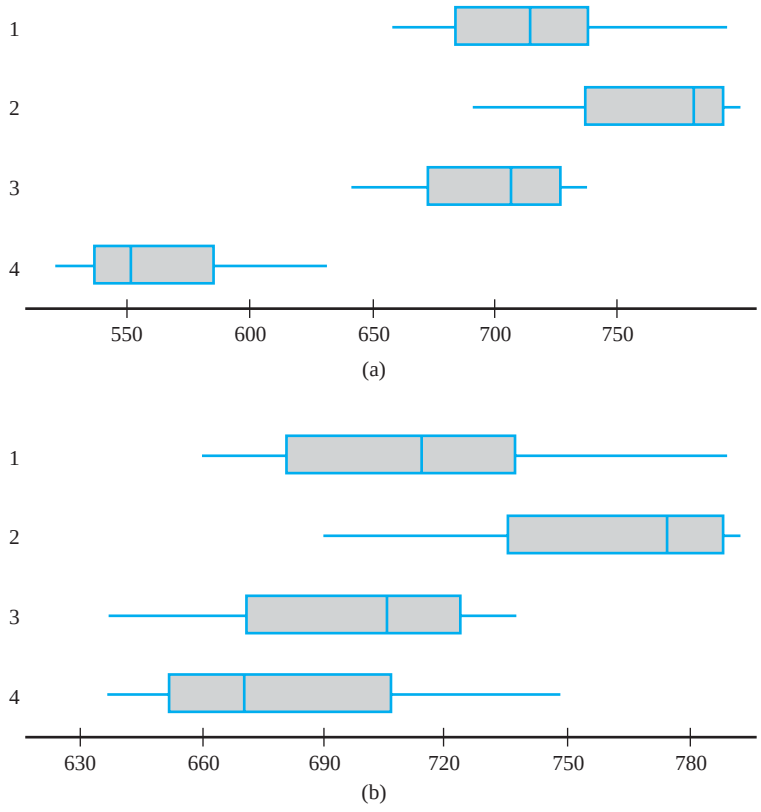


Figura 10.1 Gráficas de caja para el ejemplo 10.1: (a) datos originales; (b) datos modificados ■

Notación y suposiciones

En problemas de dos muestras se utilizaron las letras X y Y para diferenciar las observaciones en una muestra de aquéllas en la otra. Como esto es engorroso con tres o más muestras, se acostumbra utilizar una sola letra con dos subíndices. El primero identifica el número de la muestra, correspondiente a la población o tratamiento que se está muestreando y el segundo denota la posición de la observación dentro de dicha muestra. Sean

$X_{i,j}$ = la variable aleatoria (va) que denota la medición $j$ésima tomada en la población $i$ésima o la medición tomada en la $j$ésima unidad experimental que recibe el tratamiento $i$ésimo.

$x_{i,j}$ = el valor observado de $X_{i,j}$ cuando se realiza el experimento

Los datos observados normalmente se muestran en una tabla rectangular, tal como la tabla 10.1. En ella las muestras de las diferentes poblaciones aparecen en filas distintas de la tabla y $x_{i,j}$ es el número $j$ésimo en la fila $i$ésima. Por ejemplo, $x_{2,3} = 786.9$ (la tercera observación de la segunda población) y $x_{4,1} = 535.1$. Cuando no hay ambigüedad, se escribirá x_{ij} en lugar de $x_{i,j}$ (p. ej., si se realizaron 15 observaciones en cada uno de 12 tratamientos, x_{112} podría significar $x_{1,12}$ o $x_{11,2}$). Se supone que las X_{ij} dentro de cualquier muestra particular son independientes, una muestra aleatoria de la distribución de población o tratamiento $i$ésima y que las diferentes muestras son independientes entre sí.

En algunos experimentos, diferentes muestras contienen distintos números de observaciones. Aquí se abordará el caso de tamaños de muestra iguales; la generalización en cuanto a tamaños de muestra desiguales aparece en la sección 10.3. Sea J el número de observaciones en cada muestra ($J = 6$ en el ejemplo 10.1). El conjunto de datos se compone de IJ observaciones. Las medias de muestra individual serán denotadas por $\overline{X}_{1.}, \overline{X}_{2.}, \ldots, \overline{X}_{I.}$. Es decir,

$$\overline{X}_{i.} = \frac{\sum_{j=1}^{J} X_{ij}}{J} \quad i = 1, 2, \ldots, I$$

El punto en lugar del segundo subíndice significa que se sumaron todos los valores de dicho subíndice al mismo tiempo que se mantuvo fijo el valor del otro subíndice y la raya horizontal indica división entre J para obtener un promedio. Asimismo, el promedio de todas las observaciones IJ, llamada **media grande,** es

$$\overline{X}_{..} = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} X_{ij}}{IJ}$$

Con los datos en la tabla 10.1, $\overline{x}_{1.} = 713.00$, $\overline{x}_{2.} = 756.93$, $\overline{x}_{3.} = 698.07$, $\overline{x}_{4.} = 562.02$ y $\overline{x}_{..} = 682.50$. Además, sean $S_1^2, S_2^2, \ldots S_I^2$ las varianzas muestrales:

$$S_i^2 = \frac{\sum_{j=1}^{J} (X_{ij} - \overline{X}_{i.})^2}{J - 1} \quad i = 1, 2, \ldots, I$$

De acuerdo con el ejemplo 10.1, $s_1 = 46.55$, $s_1^2 = 2166.90$, y así sucesivamente.

SUPOSICIONES

Las distribuciones de población o tratamiento I son normales con la misma varianza σ^2. Es decir, cada X_{ij} está normalmente distribuida con

$$E(X_{ij}) = \mu_i \quad V(X_{ij}) = \sigma^2$$

Las desviaciones estándar de la muestra I en general difieren un poco aun cuando las σ correspondientes sean idénticas. En el ejemplo 10.1, la más grande entre s_1, s_2, s_3 y s_4 es aproximadamente 1.25 veces la más pequeña. Una regla empírica preliminar es que si la s más grande no es mucho más de dos veces la más pequeña, es razonable suponer σ^2 iguales.

En capítulos previos, se sugirió un diagrama de probabilidad normal para verificar en cuanto a normalidad. Los tamaños de muestra individuales en ANOVA típicamente son demasiado pequeños como para que los distintos diagramas I sean informativos. Se puede construir un solo diagrama restando $\bar{x}_1$ de cada observación en la primera muestra, $\bar{x}_2$ de cada observación en la segunda y así sucesivamente y luego graficando estas desviaciones IJ contra los percentiles z. La figura 10.2 da ese diagrama para los datos del ejemplo 10.1. La linealidad del patrón confirma fuertemente la suposición de normalidad.

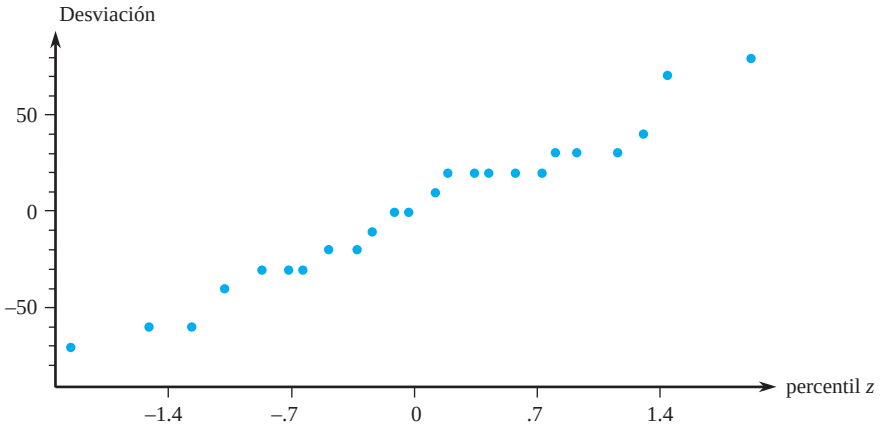


Figura 10.2 Diagrama de probabilidad normal basado en los datos del ejemplo 10.1

Si la suposición de normalidad o la suposición de varianzas iguales se supone infactible, habrá que emplear un método de análisis distinto de la prueba F usual. Búsquese por favor asesoría experta en tales situaciones (en la sección 10.3 se sugiere una posibilidad, una transformación de datos y se desarrolla otra alternativa en la sección 15.4).

El estadístico de prueba

Si H_0 es verdadera, las J observaciones en cada muestra provienen de una distribución normal con el *mismo* valor medio μ, en cuyo caso las medias muestrales $\bar{x}_{1\cdot}, \ldots, \bar{x}_{I\cdot}$ deberán ser razonablemente parecidas. El procedimiento de prueba se basa en comparar una medida de diferencias entre las $\bar{x}_{i\cdot}$ (variación "entre-muestras") con una medida de variación calculada desde *adentro* de cada una de las muestras.

DEFINICIÓN

La **media cuadrática de tratamientos** está dada, por

$$\text{MSTr} = \frac{J}{I-1}\left[(\bar{X}_{1\cdot} - \bar{X}_{\cdot\cdot})^2 + (\bar{X}_{2\cdot} - \bar{X}_{\cdot\cdot})^2 + \cdots + (\bar{X}_{I\cdot} - \bar{X}_{\cdot\cdot})^2\right]$$

$$= \frac{J}{I-1}\sum_i (\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2$$

y el **error medio cuadrático** es

$$\text{MSE} = \frac{S_1^2 + S_2^2 + \cdots + S_I^2}{I}$$

El estadístico de prueba para ANOVA unifactorial es $F = \text{MSTr/MSE}$.

La terminología "media cuadrática" se explicará en breve. Obsérvese que se utilizan $\overline{X}$ y S^2 mayúsculas, de modo que MSTr y MSE se definen como estadísticos. Se seguirá la tradición y también se utilizarán MSTr y MSE (en lugar de mstr y mse) para denotar los valores calculados de estos estadísticos. Cada S_i^2 evalúa la variación dentro de una muestra particular, así que MSE es una medida de variación dentro de muestras.

¿Qué clase de valor de F proporciona evidencia en pro o en contra de H_0? Si H_0 es verdadera (todas las μ_i son iguales), los valores de las medias muestrales individuales deberán estar próximos entre sí y por consiguiente próximos a la media grande, con el resultado de un valor relativamente pequeño de MSTr. Sin embargo, si las μ_i son bastante diferentes, algunas $\bar{x}_{i\cdot}$ difieren un poco de $\bar{x}_{\cdot\cdot}$. De modo que el valor de MSTr es afectado por la condición de H_0 (verdadera o falsa). Éste no es el caso con MSE, porque las s_i^2 dependen sólo del valor subyacente de σ^2 y no de dónde están centradas las diversas distribuciones. El siguiente recuadro presenta una propiedad importante de E(MSTr) y E(MSE), los valores esperados de estos dos estadísticos.

PROPOSICIÓN

Cuando H_0 es verdadera,

$$E(\text{MSTr}) = E(\text{MSE}) = \sigma^2$$

mientras que cuando H_0 es falsa,

$$E(\text{MSTr}) > E(\text{MSE}) = \sigma^2$$

Es decir, ambos estadísticos son insesgados para estimar la varianza de población común σ^2 cuando H_0 es verdadera, pero MSTr tiende a sobrestimar σ^2 cuando H_0 es falsa.

La insesgadez de MSE es una consecuencia de $E(S_i^2) = \sigma^2$ si H_0 es verdadera o falsa. Cuando H_0 es verdadera, cada $\overline{X}_{i\cdot}$ tiene el mismo valor medio μ y varianza σ^2/J de modo que $\sum(\overline{X}_{i\cdot} - \overline{X}_{\cdot\cdot})^2/(I-1)$, la "varianza muestral" de las $\overline{X}_{i\cdot}$, estima σ^2/J insesgadamente; multiplicando ésta por J se obtiene MSTr como un estimador insesgado de σ^2 misma. Las $\overline{X}_{i\cdot}$ tienden a dispersarse más cuando H_0 es falsa que cuando es verdadera y tiende a inflar el valor de MSTr en este caso. Por consiguiente, un valor de F que excede en gran medida 1, correspondiente a un MSTr mucho más grande que MSE, provoca una duda considerable sobre H_0. Por tanto la forma apropiada de la región de rechazo es $f \geq c$. La c de corte debe ser seleccionada para que dé $P(F \geq c$ donde H_0 es verdadera$) = \alpha$, el nivel de significancia deseado. Por consiguiente, se requiere saber la distribución de F cuando H_0 es verdadera.

Distribuciones *F* y la prueba *F*

En el capítulo 9 se introdujo una familia de distribuciones de probabilidad llamada distribuciones F en conexión con una proporción en la cual existe un número de grados de libertad (gl) asociado con el numerador y otro número de grados de libertad asociado con el denominador. Sean ν_1 y ν_2 el número de grados de libertad asociado con el numerador y denominador, respectivamente, para una variable con una distribución F. Tanto ν_1 como ν_2 son enteros positivos. La figura 10.3 ilustra una curva de densidad F y el valor crítico de cola superior correspondiente, F_{α,ν_1,ν_2}. La tabla A.9 de los apéndices da estos valores críticos para $\alpha = .10, .05, .01$ y $.001$. Los valores de ν_1 están identificados con diferentes columnas de la tabla y las filas con varios valores de ν_2. Por ejemplo, el valor crítico F que captura un área de cola superior de .05 bajo la curva F con $\nu_1 = 4$ y $\nu_2 = 6$ es $F_{.05,4,6} = 4.53$ en tanto que $F_{.05,6,4} = 6.16$. El resultado teórico clave es que el estadístico de prueba F tiene una distribución F cuando H_0 es verdadera.

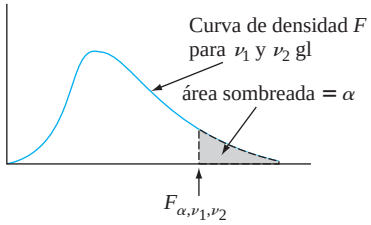


Figura 10.3 Curva de densidad F y valor crítico F_{α,ν_1,ν_2}

TEOREMA

Sea $F = \text{MSTr/MSE}$ el estadístico de prueba en un problema de ANOVA unifactorial que implica poblaciones o tratamientos I con una muestra aleatoria de J observaciones de cada uno. Cuando H_0 es verdadera y las suposiciones básicas de esta sección se satisfacen, F tiene una distribución F con $\nu_1 = I - 1$ y $\nu_2 = I(J - 1)$. Con f denotando el valor calculado de F, la región de rechazo $f \geq F_{\alpha,I-1,I(J-1)}$ especifica entonces una prueba con nivel de significancia α. Remítase a la sección 9.5 para ver cómo se obtiene información sobre el valor P para pruebas F.

El razonamiento para $\nu_1 = I - 1$ es que aunque MSTr está basada en las desviaciones I, $\overline{X}_{1\cdot} - \overline{X}_{\cdot\cdot}, \ldots, \overline{X}_{I\cdot} - \overline{X}_{\cdot\cdot}$, $\sum(\overline{X}_{i\cdot} - \overline{X}_{\cdot\cdot}) = 0$, de modo que sólo $I - 1$ de éstas son libremente determinadas. Como cada muestra contribuye con $J - 1$ grados de libertad a MSE y estas muestras son independientes, $\nu_2 = (J - 1) + \cdots + (J - 1) = I(J - 1)$.

Ejemplo 10.2 (Continuación del ejemplo 10.1)

Los valores de I y J con los datos de resistencia son 4 y 6, respectivamente, de modo que grados de libertad $= I - 1 = 3$ asociados con el numerador y grados de libertad $= I(J - 1) = 20$ asociados con el denominador. A un nivel de significancia de .05, H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ será rechazada a favor de la conclusión de que por lo menos dos μ_i son diferentes si $f \geq F_{.05,3,20} = 3.10$. La media grande es $\overline{x}_{\cdot\cdot} = \sum\sum x_{ij}/(IJ) = 682.50$,

$$\text{MSTr} = \frac{6}{4-1}[(713.00 - 682.50)^2 + (756.93 - 682.50)^2 + (698.07 - 682.50)^2 + (562.02 - 682.50)^2] = 42{,}455.86$$

$$\text{MSE} = \frac{1}{4}[(46.55)^2 + (40.34)^2 + (37.20)^2 + (39.87)^2] = 1691.92$$

$$f = \text{MSTr/MSE} = 42{,}455.86/1691.92 = 25.09$$

Como $25.09 \geq 3.10$, H_0 es resonantemente rechazada a un nivel de significancia de .05. La resistencia a la compresión promedio verdadera sí parece depender del tipo de caja. En realidad, valor P = área bajo la curva F a la derecha de $25.09 = .000$. H_0 sería rechazada a cualquier nivel de significancia razonable. ■

Ejemplo 10.3

El artículo "Influence of Contamination and Cleaning on Bond Strength to Modified Zirconia" (*Dental Materials*, 2009: 1541–1550) informó de un experimento en el que 50 discos de óxido de circonio se dividieron en cinco grupos de 10 cada uno. A continuación, un protocolo diferente de contaminación/limpieza se utilizó para cada grupo. El siguiente resumen de datos sobre la fuerza de cizalla (MPa) apareció en el artículo:

Tratamiento:	1	2	3	4	5	
Media de la muestra	10.5	14.8	15.7	16.0	21.6	Media grande = 15.7
DE muestral	4.5	6.8	6.5	6.7	6.0	

Sea μ_i que denota la verdadera fuerza de enlace promedio para el protocolo i ($i = 1, 2, 3, 4, 5$). La hipótesis nula

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

afirma que la fuerza promedio real es la misma para todos los protocolos (no depende de qué protocolo se utiliza). La hipótesis alternativa H_a establece que por lo menos dos de las μ del tratamiento son diferentes (la negación de la hipótesis nula). Los autores del artículo citado utilizaron la prueba F, así que esperaban examinar una gráfica de probabilidad normal de las desviaciones (o un gráfico diferente para cada muestra, ya que cada tamaño de la muestra es de 10) para comprobar la verosimilitud de suponer la distribución normal de la respuesta al tratamiento. La muestra de cinco desviaciones estándar son sin duda lo suficientemente cercanas entre sí para apoyar la suposición de igualdad de σ.

Los grados de libertad (gl) del numerador y denominador de la prueba son $I - 1 = 4$ e $I(J - 1) = 5(9) = 45$, respectivamente. El valor F crítico para una prueba con un nivel de significancia de .01 es $F_{.01,4,45} = 3.77$ (nuestra tabla F no tiene un grupo de renglones para 45 gl del denominador, pero la entrada de .01 para 40 gl es 3.83 y para 50 gl es 3.72). Así H_0 será rechazada si $f \geq 3.77$.

Los cuadrados de la media son

$$\begin{aligned} \text{MSTr} &= \frac{10}{5-1}[(10.5 - 15.7)^2 + (14.8 - 15.7)^2 + (15.7 - 15.7)^2 \\ &\qquad + (16.0 - 15.7)^2 + (21.6 - 15.7)^2] \\ &= 156.875 \\ \text{MSE} &= [(4.5)^2 + (6.8)^2 + (6.5)^2 + (6.7)^2 + (6.0)^2]/5 = 37.926 \end{aligned}$$

Así, el valor estadístico de prueba es $f = 156.875/37.926 = 4.14$. Este valor se encuentra en la región de rechazo ($4.14 \geq 3.77$). Al nivel de significancia .01, podemos concluir que la fuerza promedio real parece depender de qué protocolo se utiliza. Software estadístico da el valor P como .0061. ■

Cuando la hipótesis nula es rechazada por la prueba F, como ocurrió en los ejemplos 10.2 y 10.3, el experimentador suele estar interesado en el análisis posterior de los datos para decidir cuáles μ_i difieren unas de las otras. Los métodos para hacer esto se llaman *procedimientos de comparación múltiple*, que es el tema de la sección 10.2. El artículo citado en el ejemplo 10.3 resume los resultados de dicho análisis.

Sumas de los cuadrados

La introducción de las *sumas de los cuadrados* facilita el desarrollo de una apreciación intuitiva para el razonamiento que fundamenta los ANOVA unifactoriales y multifactoriales. Sea $x_{i.}$ la *suma* (no el promedio, puesto que no hay raya) de las x_{ij} con i fija (suma de los números en la $i$ésima fila de la tabla) y $x_{..}$ denota la suma de *todas* las x_{ij} (el **gran total**).

DEFINICIÓN

La **suma total de los cuadrados (SST),** la **suma de los cuadrados del tratamiento (SSTr)** y la **suma de los cuadrados del error (SSE)** están dadas por

$$\text{SST} = \sum_{i=1}^{I}\sum_{j=1}^{J}(x_{ij} - \bar{x}_{..})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}x_{ij}^2 - \frac{1}{IJ}x_{..}^2$$

$$\text{SSTr} = \sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{x}_{i.} - \bar{x}_{..})^2 = \frac{1}{J}\sum_{i=1}^{I}x_{i.}^2 - \frac{1}{IJ}x_{..}^2$$

$$\text{SSE} = \sum_{i=1}^{I}\sum_{j=1}^{J}(x_{ij} - \bar{x}_{i.})^2 \quad \text{donde } x_{i.} = \sum_{j=1}^{J}x_{ij} \quad x_{..} = \sum_{i=1}^{I}\sum_{j=1}^{J}x_{ij}$$

La suma de los cuadrados SSTr aparece en el numerador de F y SSE lo hace en el denominador de F; la razón para definir la SST se pondrá de manifiesto en breve.

Las expresiones a la extrema derecha de SST y SSTr son convenientes si los cálculos de ANOVA se realizan a mano, aunque la amplia disponibilidad de programas estadísticos hace que esto sea innecesario. Tanto SST como SSTr implican $x_{..}^2/(IJ)$ (el cuadrado del gran total dividido entre IJ), lo que normalmente se llama **factor de corrección para la media** (FC). Una vez que se calcula el factor de corrección, la SST se obtiene elevando al cuadrado cada número que aparece en la tabla, sumando los cuadrados y restando el factor de corrección. SSTr se obtiene al elevar al cuadrado cada total de fila, sumándolos, dividiendo entre J y restando el factor de corrección. SSE se calcula fácilmente como una consecuencia de la siguiente relación.

Identidad fundamental

$$\text{SST} = \text{SSTr} + \text{SSE} \tag{10.1}$$

Por consiguiente, si se calculan dos cualesquiera de las sumas de los cuadrados, la tercera se obtiene con (10.1); SST y SSTr son las más fáciles de calcular y en ese caso SSE = SST − SSTr. La comprobación se desprende de elevar al cuadrado ambos lados de la relación

$$x_{ij} - \bar{x}_{..} = (\bar{x}_{ij} - \bar{x}_{i.}) + (\bar{x}_{i.} - \bar{x}_{..}) \tag{10.2}$$

y sumando todas las i y j. Esto da la SST a la izquierda y SSTr y SSE como los dos términos extremos a la derecha. Es fácil ver que el término del producto cruz es cero.

La interpretación de la identidad fundamental es una importante ayuda para entender el ANOVA. SST mide la variación total de los datos: la suma de todas las desviaciones al cuadrado con respecto a la media grande. La identidad dice que esta variación total puede ser dividida en dos partes. SSE mide la variación que estaría presente (en las filas) aun cuando H_0 fuera verdadera o falsa y es por consiguiente la parte de la variación total que *no es explicada* por la veracidad o la falsedad de H_0. SSTr es la cantidad de variación (entre filas) que *puede ser explicada* por las posibles diferencias en las μ_i. H_0 es rechazada si la variación explicada es grande con respecto a la variación no explicada.

Una vez que SSTr y SSE se calculan, cada una se divide por su número de grados de libertad asociado para obtener un cuadrado de la media (*media* en el sentido de promedio). Entonces F es la proporción de los dos cuadrados de la media.

$$\text{MSTr} = \frac{\text{SSTr}}{I-1} \qquad \text{MSE} = \frac{\text{SSE}}{I(J-1)} \qquad F = \frac{\text{MSTr}}{\text{MSE}} \tag{10.3}$$

Con frecuencia, los cálculos se resumen en un formato tabular, llamado **tabla ANOVA,** como se ilustra en la tabla 10.2. Las tablas producidas por programas estadísticos comúnmente incluyen una columna de valor P a la derecha de f.

Tabla 10.2 Tabla ANOVA

Origen de la variación	Grados de libertad	Suma de los cuadrados	Media cuadrática	f
Tratamientos	$I-1$	SSTr	MSTr = SSTr/$(I-1)$	MSTr/MSE
Error	$I(J-1)$	SSE	MSE = SSE/$[I(J-1)]$	
Total	$IJ-1$	SST		

Ejemplo 10.4 Los datos adjuntos se obtuvieron con un experimento que compara el grado de manchado de telas copolimerizadas con tres mezclas diferentes de ácido metracrílico (datos similares aparecieron en el artículo "Chemical Factors Affecting Soiling and Soil Release from Cotton DP Fabric", *American Dyestuff Reporter*, 1983: 25–30).

						$x_{i\cdot}$	$\bar{x}_{i\cdot}$
Mezcla 1	.56	1.12	.90	1.07	.94	4.59	.918
Mezcla 2	.72	.69	.87	.78	.91	3.97	.794
Mezcla 3	.62	1.08	1.07	.99	.93	4.69	.938
						$x_{\cdot\cdot} = 13.25$	

Sea μ_i el grado promedio verdadero de manchado cuando se utiliza una mezcla i ($i = 1, 2, 3$). La hipótesis nula H_0: $\mu_1 = \mu_2 = \mu_3$ manifiesta que el grado de manchado promedio verdadero es idéntico con las tres mezclas. Se realizará una prueba a un nivel de significancia de .01 para ver si H_0 deberá ser rechazada a favor de la aseveración de que el grado de manchado promedio verdadero no es el mismo con todas las mezclas. Como $I - 1 = 2$ e $I(J - 1) = 12$, H_0 deberá ser rechazada si $f \geq F_{.01,2,12} = 6.93$. Elevando al cuadrado cada una de las 15 observaciones y sumando se obtiene $\sum\sum x_{ij}^2 = (.56)^2 + (1.12)^2 + \cdots + (.93)^2 = 12.1351$. Los valores de las tres sumas de los cuadrados son

$$\begin{aligned} \text{SST} &= 12.1351 - (13.25)^2/15 = 12.1351 - 11.7042 = .4309 \\ \text{SSTr} &= \frac{1}{5}[(4.59)^2 + (3.97)^2 + (4.69)^2] - 11.7042 \\ &= 11.7650 - 11.7042 = .0608 \\ \text{SSE} &= .4309 - .0608 = .3701 \end{aligned}$$

El resto de los cálculos se ilustra en la tabla ANOVA adjunta. Como $f = .99 < 6.93$, H_0 no es rechazada a un nivel de significancia de .01. Parece que las mezclas son indistinguibles con respecto al grado de manchado ($F_{.10,2,12} = 2.81 \Rightarrow$ valor $P > .10$).

Origen de la variación	Grados de libertad	Suma de los cuadrados	Cuadrado de la media	f
Tratamientos	2	.0608	.0304	.99
Error	12	.3701	.0308	
Total	14	.4309		

EJERCICIOS Sección 10.1 (1–10)

1. En un experimento para comparar las resistencias a la tensión de $I = 5$ tipos diferentes de alambre de cobre, se utilizaron $J = 4$ muestras de cada tipo. Las estimaciones entre muestras y dentro de muestras de σ^2 se calcularon como MSTr = 2673.3 y MSE = 1094.2, respectivamente.

a. Use la prueba F a un nivel de .05 para probar H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ contra H_a: por lo menos dos μ_i son desiguales.

b. ¿Qué se puede decir sobre el valor P para la prueba?

2. Suponga que las observaciones de resistencia a la compresión del cuarto tipo de caja del ejemplo 10.1 hubieran sido 655.1, 748.7, 662.4, 679.0, 706.9 y 640.0 (obtenidas sumando 120 a cada x_{4j} previa). Suponiendo que las observaciones restantes no cambian, realice una prueba F con $\alpha = .05$.

3. Se determinó el rendimiento en lúmenes de cada uno de $I = 3$ marcas diferentes de focos de luz blanca de 60 watts, con $J = 8$ focos de cada marca probados. Las sumas de los cuadrados se

calcularon como SSE = 4773.3 y SSTr = 591.2. Formule las hipótesis de interés (incluidas definiciones en palabras de los parámetros) y use la prueba F de ANOVA ($\alpha = .05$) para decidir si existen diferencias en los rendimientos de lúmenes promedio verdaderos entre las tres marcas de este tipo de foco obteniendo tanta información como sea posible sobre el valor P.

4. Es una práctica común en muchos países destruir (fragmentar) refrigeradores al final de su vida útil. En este proceso el material de espuma de aislamiento puede ser liberado a la atmósfera. El artículo "Release of Fluorocarbons from Insulation Foam in Home Appliances during Shredding" (*J. of the Air and Waste Mgmt. Assoc.*, 2007: 1452–1460) dio los siguientes datos sobre la densidad de la espuma (g/L) para cada uno de dos refrigeradores producidos por cuatro distintos fabricantes:

1. 30.4, 29.2 2. 27.7, 27.1
3. 27.1, 24.8 4. 25.5, 28.8

¿Parece que el promedio real de densidad de la espuma no es el mismo para todos estos fabricantes? Lleve a cabo una prueba adecuada de las hipótesis mediante la obtención de una mayor cantidad de información del valor P como sea posible y un resumen de su análisis en una tabla de ANOVA.

5. Considere los siguientes datos del módulo de elasticidad ($\times 10^6$ lb/pulg^2) de madera de tres grados diferentes (en concordancia con los valores que aparecen en el artículo "Bending Strength and Stiffness of Second-Growth Douglas-Fir Dimension Lumber" (*Forest Products J.*, 1991: 35–43), excepto que los tamaños de muestra allí eran más grandes):

Grado	J	$\bar{x}_{i\cdot}$	s_i
1	10	1.63	.27
2	10	1.56	.24
3	10	1.42	.26

Use estos datos y un nivel de significancia de .01 para probar la hipótesis nula de no diferencia en el módulo medio de elasticidad para los tres grados.

6. El artículo "Origin of Precambrian Iron Formations" (*Econ. Geology*, 1964: 1025-1057) reporta los siguientes datos sobre Fe total para cuatro tipos de formación de hierro (1 = carbonato, 2 = silicato, 3 = magnetita, 4 = hematita).

1:	20.5	28.1	27.8	27.0	28.0
	25.2	25.3	27.1	20.5	31.3
2:	26.3	24.0	26.2	20.2	23.7
	34.0	17.1	26.8	23.7	24.9
3:	29.5	34.0	27.5	29.4	27.9
	26.2	29.9	29.5	30.0	35.6
4:	36.5	44.2	34.1	30.3	31.4
	33.1	34.1	32.9	36.3	25.5

Analice una prueba F de varianza a un nivel de significancia de .01 y resuma los resultados en una tabla ANOVA.

7. Se realizó un experimento para comparar la resistencia eléctrica de seis diferentes mezclas de hormigón de baja permeabilidad cubierta del puente. Hubo 26 mediciones en cilindros de concreto para cada mezcla, los cuales se obtuvieron 28 días después de la fundición. Las entradas de la tabla de ANOVA de acompañamiento se basan en la información en el artículo "In-Place Resistivity of Bridge Deck Concrete Mixtures" (ACI Materials J., 2009: 114–122). Rellene el resto de las entradas y la prueba de hipótesis adecuada.

Fuente	Grados de libertad	Suma de los cuadrados	Media cuadrática	f
Mezcla				
Error			13.929	
Total		5664.415		

8. Un estudio de las propiedades de armaduras conectadas con placas metálicas para soportar techos ("Modeling Joints Made with Light-Gauge Metal Connector Plates", *Forest Products J.*, 1979: 39–44) dio las siguientes observaciones de índice de rigidez axial (kips/pulg) de tramos de placa de 4, 6, 8, 10 y 12 pulg:

4:	309.2	409.5	311.0	326.5	316.8	349.8	309.7
6:	402.1	347.2	361.0	404.5	331.0	348.9	381.7
8:	392.4	366.2	351.0	357.1	409.9	367.3	382.0
10:	346.7	452.9	461.4	433.1	410.6	384.2	362.6
12:	407.4	441.8	419.9	410.7	473.4	441.2	465.8

¿Tiene algún efecto la variación de la longitud de placas en la rigidez axial promedio verdadera? Formule y pruebe las hipótesis pertinentes mediante un análisis de varianza con $\alpha = .01$. Muestre sus resultados en una tabla ANOVA. [*Sugerencia*: $\sum\sum x_{ij}^2 = 5{,}241{,}420.79$.]

9. Se analizaron seis muestras de cada uno de cuatro tipos de crecimiento de granos de cereal en una región para determinar el contenido de tiamina y se obtuvieron los siguientes resultados (μg/g):

Trigo	5.2	4.5	6.0	6.1	6.7	5.8
Cebada	6.5	8.0	6.1	7.5	5.9	5.6
Maíz	5.8	4.7	6.4	4.9	6.0	5.2
Avena	8.3	6.1	7.8	7.0	5.5	7.2

¿Sugieren estos datos que por lo menos dos de los granos difieren con respecto al contenido de tiamina promedio verdadero? Use un nivel $\alpha = .05$ con base en el método del valor P.

10. En ANOVA unifactorial con tratamientos I y observaciones J por cada tratamiento, sea $\mu = (1/I)\sum\mu_i$.

a. Exprese $E(\bar{X}_{\cdot})$ en función de μ. [*Sugerencia*: $\bar{X}_{\cdot\cdot} = (1/I)\sum\bar{X}_{i\cdot}$.]

b. Calcule $E(\bar{X}_{i\cdot}^2)$. [*Sugerencia*: con cualquier variable aleatoria Y, $E(Y^2) = V(Y) + [E(Y)]^2$.]

c. Calcule $E(\bar{X}_{\cdot}^2)$.

d. Calcule $E(\text{SSTr})$ y luego demuestre que

$$E(\text{MSTr}) = \sigma^2 + \frac{J}{I-1}\sum(\mu_i - \mu)^2$$

e. Con el resultado del inciso (d), ¿cuál es $E(\text{MSTr})$ cuando H_0 es verdadera? Cuando H_0 es falsa, ¿se compara $E(\text{MSTr})$ con σ^2?

10.2 Comparaciones múltiples en ANOVA

Cuando el valor calculado del estadístico F en un ANOVA unifactorial no es significativo, el análisis se termina porque no se han identificado diferencias entre las μ_i. Pero cuando H_0 es rechazada, el investigador normalmente deseará saber cuáles de las μ_i son diferentes una de otra. Un método para realizar este análisis adicional se llama **procedimiento de comparaciones múltiples.**

Varios de dichos procedimientos más frecuentemente utilizados están basados en la siguiente idea central. Primero se calcula un intervalo de confianza para cada diferencia $\mu_i - \mu_j$ con $i < j$. Por consiguiente si $I = 4$, los seis intervalos de confianza requeridos serían para $\mu_1 - \mu_2$ (pero no también para $\mu_2 - \mu_1$), $\mu_1 - \mu_3$, $\mu_1 - \mu_4$, $\mu_2 - \mu_3$, $\mu_2 - \mu_4$ y $\mu_3 - \mu_4$. Entonces si el intervalo para $\mu_1 - \mu_2$ no incluye 0, se concluye que μ_1 y μ_2 *difieren significativamente* una de otra, si el intervalo sí incluye 0, se considera que las dos μ no difieren de manera significativa. Si se sigue la misma línea de razonamiento para cada uno de los demás intervalos, finalmente se es capaz de juzgar si cada par de μ difiere o no en forma significativa una de otra.

Los procedimientos basados en esta idea difieren en el método utilizado para calcular los varios intervalos de confianza. Aquí se presenta un método popular que controla el nivel de confianza *simultáneo* para todos los intervalos $I(I-1)/2$.

Procedimiento de Tukey (el método *T*)

El procedimiento de Tukey implica utilizar otra distribución de probabilidad llamada **distribución de rango estudentizado.** La distribución depende de dos parámetros: m grados de libertad asociados con el numerador y grados de libertad asociados con el denominador. Sea $Q_{\alpha,m,\nu}$ el valor crítico α de cola superior de la distribución de rango estudentizado con m grados de libertad asociados con el numerador y ν grados de libertad asociados con el denominador (análogo a F_{α,ν_1,ν_2}). En la tabla A.10 del Apéndice se dan valores de $Q_{\alpha,m,\nu}$.

PROPOSICIÓN

Con la probabilidad $1 - \alpha$,

$$\bar{X}_{i\cdot} - \bar{X}_{j\cdot} - Q_{\alpha,I,I(J-1)}\sqrt{\text{MSE}/J} \le \mu_i - \mu_j \le \bar{X}_{i\cdot} - \bar{X}_{j\cdot} + Q_{\alpha,I,I(J-1)}\sqrt{\text{MSE}/J} \qquad (10.4)$$

para *cada* i y j ($i = 1, \ldots, I$ y $j = 1, \ldots, I$) con $i < j$.

Obsérvese que los grados de libertad asociados con el numerador para el valor crítico Q_α apropiado es I, el número de medias de la población o tratamiento que se están comparando y no $I - 1$ como en la prueba F. Cuando las $\bar{x}_{i\cdot}$, $\bar{x}_{j\cdot}$ son calculadas y MSE se sustituyen en (10.4), el resultado es un conjunto de intervalos de confianza con nivel de confianza *simultáneo* de $100(1-\alpha)\%$ para todas las diferencias de la forma $\mu_i - \mu_j$ con $i < j$. Cada intervalo que no incluye 0 da lugar a la conclusión de que los valores correspondientes de μ_i y μ_j difieren significativamente uno de otro.

Como en realidad no interesan los límites inferior y superior de los diversos intervalos sino sólo cuál incluye 0 y cuál no, se puede evitar mucha de la aritmética asociada con (10.4). El siguiente recuadro da detalles y describe cómo se pueden identificar las diferencias de modo visual con un “patrón de subrayado”.

Método *T* para identificar μ_i significativamente diferentes

Se selecciona α, se extrae $Q_{\alpha,I,I(J-1)}$ de la tabla A.10 del Apéndice y se calcula $w = Q_{\alpha,I,I(J-1)} \cdot \sqrt{\text{MSE}/J}$. Luego se hace una lista de las medias muestrales en orden creciente y se subrayan los pares que difieren menos de w. Cualquier par de medias muestrales no subrayado por la misma raya corresponde a un par de medias de población o tratamiento juzgadas significativamente diferentes.

Supóngase, por ejemplo, que $I = 5$ y que

$$\bar{x}_{2\cdot} < \bar{x}_{5\cdot} < \bar{x}_{4\cdot} < \bar{x}_{1\cdot} < \bar{x}_{3\cdot}$$

Entonces

1. Considere en primer lugar la media más pequeña $\bar{x}_{2\cdot}$. Si $\bar{x}_{5\cdot} - \bar{x}_{2\cdot} \geq w$, prosiga al paso 2. Sin embargo, si $\bar{x}_{5\cdot} - \bar{x}_{2\cdot} < w$, conecte estas primeras dos medias con un segmento de línea. Luego si es posible extienda este segmento de recta más a la derecha de la $\bar{x}_{i\cdot}$ más grande que difiera de $\bar{x}_{2\cdot}$ en menos de w (de modo que la recta pueda conectar dos, tres o incluso más medias).
2. Ahora siga con $\bar{x}_{5\cdot}$ y otra vez extienda el segmento de línea hasta la derecha de la $\bar{x}_{i\cdot}$ más grande que difiera de $\bar{x}_{5\cdot}$ en menos de w (puede que no sea posible trazar esta línea o alternativamente puede que subraye sólo dos medias o tres o incluso las cuatro medias restantes).
3. Continúe con $\bar{x}_{4\cdot}$ y repita y finalmente continúe con $\bar{x}_{1\cdot}$.

Para resumir, comenzando en cada media que aparece en la lista ordenada, un segmento de recta se extiende tan lejos a la derecha como sea posible en tanto que la diferencia entre las medias sea menor que w. Es fácil verificar que un intervalo particular de la forma (10.4) contendrá 0 si y sólo si el par correspondiente de medias muestrales está subrayado por el mismo segmento de recta.

Ejemplo 10.5 Se realizó un experimento para comparar cinco marcas diferentes de filtros de aceite para automóviles con respecto a su capacidad de atrapar materia extraña. Sea μ_i la cantidad promedio verdadera de material atrapado por los filtros marca i ($i = 1, \ldots, 5$) en condiciones controladas. Se utilizó una muestra de nueve filtros de cada marca y se obtuvieron las siguientes cantidades medias muestrales: $\bar{x}_{1\cdot} = 14.5$, $\bar{x}_{2\cdot} = 13.8$, $\bar{x}_{3\cdot} = 13.3$, $\bar{x}_{4\cdot} = 14.3$ y $\bar{x}_{5\cdot} = 13.1$. La tabla 10.3 es una tabla ANOVA que resume la primera parte del análisis.

Tabla 10.3 Tabla ANOVA para el ejemplo 10.5

Origen de la variación	Grados de libertad	Suma de los cuadrados	Media cuadrática	f
Tratamientos (marcas)	4	13.32	3.33	37.84
Error	40	3.53	.088	
Total	44	16.85		

Como $F_{.05,4,40} = 2.61$, H_0 es rechazada (decisivamente) a un nivel de .05. Ahora utilice el procedimiento de Tukey para buscar diferencias significativas entre las μ_i. En la tabla A.10 del apéndice, $Q_{.05,5,40} = 4.04$ (el segundo subíndice de Q es I y no $I - 1$ como en F), por lo tanto $w = 4.04\sqrt{.088/9} = .4$. Después de ordenar las cinco medias muestrales en orden creciente, un segmento de línea puede conectar a la dos más pequeñas porque difieren por

menos de .4. No obstante, este segmento no puede ser extendido más a la derecha puesto que $13.8 - 13.1 = .7 \geq .4$. Moviéndose una media a la derecha, el par $\bar{x}_{3.}$ y $\bar{x}_{2.}$ no puede ser subrayado porque estas medias difieren por más de .4. De nuevo moviéndose a la derecha, la siguiente media, 13.8, no puede ser conectada a algo que esté más a la derecha. Las dos últimas medias pueden ser subrayadas con el mismo segmento de línea.

$\bar{x}_{5.}$	$\bar{x}_{3.}$	$\bar{x}_{2.}$	$\bar{x}_{4.}$	$\bar{x}_{1.}$
13.1	13.3	13.8	14.3	14.5

Así pues las marcas 1 y 4 no son significativamente diferentes una de otra, pero sí son más altas de manera significativa que las otras tres marcas en sus contenidos promedio verdaderos. La marca 2 es significativamente mejor que la 3 y 5 pero peor que la 1 y 4 y las marcas 3 y 5 no difieren en modo significativo.

Si $\bar{x}_{2.} = 14.15$ en lugar de 13.8 con el mismo valor w calculado, entonces la configuración de medias subrayadas sería

$\bar{x}_{5.}$	$\bar{x}_{3.}$	$\bar{x}_{2.}$	$\bar{x}_{4.}$	$\bar{x}_{1.}$
13.1	13.3	14.15	14.3	14.5

■

Ejemplo 10.6 Un biólogo deseaba estudiar los efectos del etanol en el periodo de sueño. Se seleccionó una muestra de 20 ratas equiparadas por edad y otras características y a cada rata se le administró una inyección oral con una concentración particular de etanol por peso corporal. Luego se registró el periodo de sueño de movimiento rápido de ojos (REM, por sus siglas en inglés) de cada rata durante 24 horas, con los siguientes resultados:

Tratamiento (concentración de etanol)						$x_{i.}$	$\bar{x}_{i.}$
0 (control)	88.6	73.2	91.4	68.0	75.2	396.4	79.28
1 g/kg	63.0	53.9	69.2	50.1	71.5	307.7	61.54
2 g/kg	44.9	59.5	40.2	56.3	38.7	239.6	47.92
4 g/kg	31.0	39.6	45.3	25.2	22.7	163.8	32.76

$x_{..} = 1107.5$ $\bar{x}_{..} = 55.375$

¿Indican los datos que el promedio real del periodo de sueño REM depende de la concentración de etanol? (Este ejemplo está basado en el experimento reportado en "Relationship of Ethanol Blood Level to REM and Non-REM Sleep Time and Distribution in the Rat", *Life Sciences*, 1978: 839–846.)

Las $\bar{x}_{i.}$ difieren sustancialmente una de otra, aunque también existe una gran cantidad de variabilidad dentro de cada muestra, por lo que para responder la pregunta con precisión se debe realizar el ANOVA. Con $\sum\sum x_{ij}^2 = 68{,}697.6$ y el factor de corrección $x_{..}^2/(IJ) = (1107.5)^2/20 = 61{,}327.8$, las fórmulas calculadas dan

$$\begin{aligned} \text{SST} &= 68{,}697.6 - 61{,}327.8 = 7369.8 \\ \text{SSTr} &= \frac{1}{5}[(396.40)^2 + (307.70)^2 + (239.60)^2 + (163.80)^2] - 61{,}327.8 \\ &= 67{,}210.2 - 61{,}327.8 = 5882.4 \\ \text{SSE} &= 7369.8 - 5882.4 = 1487.4 \end{aligned}$$

La tabla 10.4 es una tabla ANOVA SAS. La última columna da el valor P como .0001. Con un nivel de significancia de .05, se rechaza la hipótesis nula H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$, puesto que valor $P = .0001 < .05 = \alpha$. Parece que el promedio real del periodo de sueño REM depende del nivel de concentración.

Tabla 10.4 Tabla ANOVA SAS

Análisis de procedimientos de varianza
Variable dependiente: TIEMPO

Fuente	DF	Suma de cuadrados	Media cuadrática	Valor F	Pr > F
Modelo	3	5882.35750	1960.78583	21.09	0.0001
Error	16	1487.40000	92.96250		
corregido					
Total	19	7369.75750			

Existen $I = 4$ tratamientos y 15 grados de libertad asociados con el error, por lo tanto $Q_{.05,4,16} = 4.05$ y $w = 4.05\sqrt{93.0/5} = 17.47$. Ordenando las medias y subrayándolas se obtiene

$\bar{x}_{4.}$	$\bar{x}_{3.}$	$\bar{x}_{2.}$	$\bar{x}_{1.}$
32.76	47.92	61.54	79.28

La interpretación de este subrayado debe hacerse con cuidado, puesto que parece que se ha concluido que los tratamientos 2 y 3 no difieren, 3 y 4 no difieren y no obstante 2 y 4 sí lo hacen. La forma sugerida de expresar esto es decir que aunque la evidencia permite concluir que los tratamientos 2 y 4 difieren uno de otro, no se ha demostrado que alguno es significativamente diferente del 3. El tratamiento 1 tiene un periodo de sueño REM promedio verdadero más alto de manera significativa que cualquiera de los demás tratamientos.

La figura 10.4 muestra resultados obtenidos con SAS a partir de la aplicación del procedimiento de Tukey.

Alfa = 0.05 gl = 16 MSE = 92.9625
Valor crítico de rango estudentizado = 4.046
Diferencia significativa mínima = 17.446

Medias con la misma letra no son significativamente diferentes.

Agrupamiento de Tukey	Media	N	TRATAMIENTO
A	79.280	5	0(control)
B	61.540	5	1 gm/kg
B			
C B	47.920	5	2 gm/kg
C			
C	32.760	5	4 gm/kg

Figura 10.4 Método de Turkey usando SAS ■

Interpretación de α en el método de Tukey

Previamente se manifestó que el método de Tukey controla el nivel de confianza *simultáneo*. Entonces ¿qué significa "simultáneo" en este caso? Calcúlese un intervalo de confianza de 95% para una media de población μ basada en una muestra de dicha población y luego un intervalo de confianza de 95% para una proporción de población p basado en otra muestra seleccionada independientemente de la primera. Antes de obtener los datos, la probabilidad de que el primer intervalo incluya μ es de .95 y ésta también es la probabilidad de que el segundo intervalo incluya p. Como las dos muestras se seleccionan de manera independiente una de otra, la probabilidad de que *ambos* intervalos incluyan los valores de los parámetros respectivos es $(.95)(.95) = (.95)^2 \approx .90$. Por consiguiente, el nivel de confianza *simultáneo* o *conjunto* para los dos intervalos es aproximadamente de 90%, si se calculan pares de intervalos una y otra vez con muestras independientes, a la

larga aproximadamente 90% de las veces el primer intervalo capturará μ y el segundo incluirá p. Asimismo, si se calculan tres intervalos de confianza basados en muestras independientes, el nivel de confianza simultáneo será de $100(.95)^3\% \approx 86\%$. Claramente, a medida que se incrementa el número de intervalos, el nivel de confianza simultáneo de que todos los intervalos capturen sus respectivos parámetros se reducirá.

Ahora supóngase que se desea mantener el nivel de confianza simultáneo en 95%. Entonces para dos muestras independientes, el nivel de confianza individual para cada una tendría que ser de $100\sqrt{.95}\% \approx 97.5\%$. Mientras más grande es el número de intervalos, más alto tendría que ser el nivel de confianza individual para mantener el nivel simultáneo en 95%.

El truco en relación con los intervalos Tukey es que no están basados en muestras independientes, MSE aparece en todos y varios intervalos comparten las mismas $\bar{x}_{i\cdot}$' (p. ej., en el caso $I = 4$, tres intervalos diferentes utilizan $\bar{x}_{1\cdot}$). Esto implica que no existe un argumento de probabilidad directo para discernir el nivel de confianza simultáneo de los niveles de confianza individuales. No obstante, se puede demostrar que si se utiliza $Q_{.05}$, el nivel de confianza simultáneo se controla a 95%, en tanto que si se utiliza $Q_{.01}$ se obtiene un nivel simultáneo de 99%. Para obtener un nivel simultáneo de 95%, el nivel individual de cada intervalo debe ser considerablemente más grande que 95%. Expresado en una forma un poco diferente, para obtener una proporción de error de 5% asociada con un *experimento* o *familia*, la proporción de error por comparación o individual para cada intervalo debe ser considerablemente más pequeña que .05. Minitab le pide al usuario que especifique la proporción de error asociado con la familia (p. ej., 5%) y luego incluye en los datos de salida la proporción de error individual (véase el ejercicio 16).

Intervalos de confianza para otras funciones paramétricas

En algunas situaciones, se desea un intervalo de confianza para una función de las μ_i más complicada que una diferencia $\mu_i - \mu_j$. Sea $\theta = \sum c_i\mu_i$, donde las c_i son constantes. Una función como ésa es $\frac{1}{2}(\mu_1 + \mu_2) - \frac{1}{3}(\mu_3 + \mu_4 + \mu_5)$, la cual en el contexto del ejemplo 10.5 mide la diferencia entre el grupo compuesto de las dos primeras marcas y la de las últimas tres. Como las X_{ij} están normalmente distribuidas con $E(X_{ij}) = \mu_i$ y $V(X_{ij}) = \sigma^2$, $\hat{\theta} = \sum c_i\bar{X}_i$ está normalmente distribuida, insesgada para θ, y

$$V(\hat{\theta}) = V\left(\sum_i c_i\bar{X}_{i\cdot}\right) = \sum_i c_i^2 V(\bar{X}_{i\cdot}) = \frac{\sigma^2}{J}\sum_i c_i^2$$

La estimación de σ^2 mediante MSE y la formación de $\hat{\sigma}_{\hat{\theta}}$ da por resultado una variable t $(\hat{\theta} - \theta)/\hat{\sigma}_{\hat{\theta}}$, la cual puede ser manipulada para obtener el siguiente intervalo de confianza de $100(1 - \alpha)\%$ para $\sum c_i\mu_i$.

$$\sum c_i\bar{x}_{i\cdot} \pm t_{\alpha/2,I(J-1)}\sqrt{\frac{\text{MSE}\sum c_i^2}{J}} \tag{10.5}$$

Ejemplo 10.7 La función paramétrica para comparar las primeras dos marcas de filtro de aceite (tienda) con las últimas tres marcas (nacionales) es $\theta = \frac{1}{2}(\mu_1 + \mu_2) - \frac{1}{3}(\mu_3 + \mu_4 + \mu_5)$, con la cual

$$\sum c_i^2 = \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 + \left(-\frac{1}{3}\right)^2 + \left(-\frac{1}{3}\right)^2 + \left(-\frac{1}{3}\right)^2 = \frac{5}{6}$$

$$\text{Con } \hat{\theta} = \frac{1}{2}(\bar{x}_{1\cdot} + \bar{x}_{2\cdot}) - \frac{1}{3}(\bar{x}_{3\cdot} + \bar{x}_{4\cdot} + \bar{x}_{5\cdot}) = .583 \text{ y MSE} = .088, \text{ un intervalo de 95\% es}$$

$$.583 \pm 2.021\sqrt{5(.088)/[(6)(9)]} = .583 \pm .182 = (.401, .765)$$ ■

En ocasiones se realiza un experimento para comparar cada uno de varios tratamientos "nuevos" con un tratamiento de control. En tales situaciones, una técnica de comparaciones múltiples llamada método de Dunnett es apropiada.

EJERCICIOS Sección 10.2 (11–21)

11. En un experimento para comparar las proporciones de cobertura de cinco marcas diferentes de pintura amarilla de látex para interiores disponibles en un área particular se utilizaron 4 galones ($J = 4$) de cada pintura. Las proporciones de cobertura promedio de las muestras (pies^2/gal) de las cinco marcas fueron $\bar{x}_{1\cdot} = 462.0$, $\bar{x}_{2\cdot} = 512.8$, $\bar{x}_{3\cdot} = 437.5$, $\bar{x}_{4\cdot} = 469.3$ y $\bar{x}_{5\cdot} = 532.1$. Se encontró que el valor calculado de F es significativo a un nivel de $\alpha = .05$. Con MSE = 272.8 use el procedimiento de Tukey para investigar diferencias significativas en las proporciones de cobertura promedio verdaderas entre marcas.

12. En el ejercicio 11, suponga $\bar{x}_{3\cdot} = 427.5$. Ahora, ¿cuáles proporciones de cobertura promedio verdaderas difieren significativamente una de otra? Asegúrese de utilizar el método de subrayar para ilustrar sus conclusiones y escriba un párrafo que resuma sus resultados.

13. Repita el ejercicio 12 suponiendo que $\bar{x}_{2\cdot} = 502.8$ además de $\bar{x}_{3\cdot} = 427.5$.

14. Use el procedimiento de Tukey con los datos del ejemplo 10.3 para identificar diferencias en la fuerza adhesiva real promedio entre los cinco protocolos.

15. El ejercicio 10.7 describe un experimento en el que se realizaron 26 observaciones de resistencia en cada una de seis diferentes mezclas de concreto. El artículo citado dio las siguientes medias muestrales: 14.18, 17.94, 18.00, 18.00, 25.74, 27.67. Aplique el método de Tukey con un nivel de confianza simultáneo de 95% para identificar diferencias significativas, así como de sus resultados (use MSE = 13.929).

16. Reconsidere los datos de rigidez axial dados en el ejercicio 8. Los siguientes son datos ANOVA obtenidos con Minitab.

```
Análisis de varianza para rígidez
Fuente    DF       SS       MS        F        P
Longitud   4    43993    10998    10.48    0.000
Error     30    31475     1049
Total     34    75468

Nivel     N     Mean   StDev
4         7   333.21   36.59
6         7   368.06   28.57
8         7   375.13   20.83
10        7   407.36   44.51
12        7   437.17   26.00

DE agrupada = 32.39

Comparaciones de pares de Tukey

Rapidez de error de la familia  = 0.0500
   Rapidez de error individual  = 0.00693

Valor crítico  = 4.10

Intervalos para
(Media a nivel columna) – (Media a nivel renglón)
          4          6          8         10

6       −85.0
         15.4

8       −92.1      −57.3
          8.3       43.1

10     −124.3      −89.5      −82.4
         −23.9       10.9       18.0

12     −154.2     −119.3     −112.2      −80.0
         −53.8      −18.9      −11.8       20.4
```

a. ¿Es factible que las varianzas de las cinco distribuciones de índices de rigidez axial sean idénticas? Explique.

b. Use los resultados (sin referencia a la tabla F) para probar las hipótesis pertinentes.

c. Use los intervalos de Tukey dados en los resultados para determinar cuáles medias difieren y construya el patrón de subrayado correspondiente.

17. Remítase al ejercicio 5. Calcule un intervalo de confianza t de 95% con $\theta = \frac{1}{2}(\mu_1 + \mu_2) - \mu_3$.

18. Considere los datos adjuntos sobre crecimiento de plantas después de la aplicación de cinco diferentes tipos de la hormona del crecimiento.

1:	13	17	7	14
2:	21	13	20	17
3:	18	15	20	17
4:	7	11	18	10
5:	6	11	15	8

a. Realice una prueba F al nivel $\alpha = .05$.

b. ¿Qué sucede cuando se aplica el procedimiento de Tukey?

19. Considere un experimento ANOVA unifactorial en el cual $I = 3$, $J = 5$, $\bar{x}_{1\cdot} = 10$, $\bar{x}_{2\cdot} = 12$ y $\bar{x}_{3\cdot} = 20$. Encuentre un valor de SSE con el cual $f > F_{.05,2,12}$ de modo que H_0: $\mu_1 = \mu_2 = \mu_3$ sea rechazada, aunque cuando se aplica el procedimiento de Tukey se puede decir que ninguna de las μ_i difieren significativamente una de otra.

20. Remítase al ejercicio 19 y suponga $\bar{x}_{1\cdot} = 10, \bar{x}_{2\cdot} = 15$ y $\bar{x}_{3\cdot} = 20$. ¿Puede hallar ahora un valor de SSE que produzca semejante contradicción entre la prueba F y el procedimiento de Tukey?

21. El artículo "The Effect of Enzyme Inducing Agents on the Survival Times of Rats Exposed to Lethal Levels of Nitrogen Dioxide" (*Toxicology and Applied Pharmacology*, 1978: 169–174) reporta los siguientes datos sobre tiempos de sobrevivencia de ratas expuestas a bióxido de nitrógeno (70 ppm) vía diferentes regímenes de inyección. Hubo $J = 14$ ratas en cada grupo.

Régimen	$\bar{x}_{i\cdot}$ (min)	s_i
1. Control	166	32
2. 3-Metilcolantreno	303	53
3. Alilisopropilacetamida	266	54
4. Fenobarbital	212	35
5. Cloropromazina	202	34
6. Ácido *p*-Aminobenzoico	184	31

a. Pruebe las hipótesis nulas de que el tiempo de sobrevivencia promedio verdadero no depende del régimen de inyección contra la alternativa de que existe alguna dependencia del régimen de inyección con $\alpha = .01$.

b. Suponga que se calculan intervalos de confianza de $100(1 - \alpha)\%$ para k funciones paramétricas diferentes con el mismo conjunto de datos ANOVA. Entonces es fácil verificar que el nivel de confianza simultáneo es por lo menos de $100(1 - k\alpha)\%$. Calcule intervalos de confianza con nivel de confianza simultáneo de por lo menos 98% para $\mu_1 - \frac{1}{5}(\mu_2 + \mu_3 + \mu_4 + \mu_5 + \mu_6)$ y $\frac{1}{4}(\mu_2 + \mu_3 + \mu_4 + \mu_5) - \mu_6$

10.3 Más sobre ANOVA unifactorial

A continuación se consideran con brevedad algunos temas adicionales relacionados con ANOVA unifactorial. Éstos incluyen una descripción alternativa de los parámetros modelo, β para la prueba F, la relación de la prueba con los procedimientos previamente considerados, la transformación de datos, un modelo de efectos aleatorios y las fórmulas para el caso de tamaños de muestra desiguales.

El modelo ANOVA

Las suposiciones de ANOVA unifactorial pueden ser descritas sucintamente por medio de la "ecuación modelo"

$$X_{ij} = \mu_i + \epsilon_{ij}$$

donde ϵ_{ij} representa una desviación aleatoria de la población o de la media de tratamiento verdadera μ_i. Se supone que las ϵ_{ij} son variables aleatorias independientes normalmente distribuidas (lo que implica que las X_{ij} también lo son) con $E(\epsilon_{ij}) = 0$ [de modo que $E(X_{ij}) = \mu_i$] y $V(\epsilon_{ij}) = \sigma^2$ [de donde $V(X_{ij}) = \sigma^2$ para toda i y j]. Una descripción alternativa de ANOVA unifactorial dará una idea adicional y sugerirá generalizaciones apropiadas de modelos que implican más de un factor. Defínase el parámetro μ como

$$\mu = \frac{1}{I}\sum_{i=1}^{I}\mu_i$$

y los parámetros $\alpha_1, \ldots, \alpha_I$ como

$$\alpha_i = \mu_i - \mu \quad (i = 1, \ldots, I)$$

Entonces la media del tratamiento μ_i se escribe como $\mu + \alpha_i$, donde μ representa la respuesta total promedio verdadera en el experimento y α_i es el efecto, medido como un alejamiento de μ, debido al $i$ésimo tratamiento. Mientras que inicialmente se tenían I parámetros, ahora se tienen $I + 1$ parámetros ($\mu, \alpha_1, \ldots, \alpha_I$). Sin embargo, como $\sum\alpha_i = 0$ (el alejamiento promedio de la respuesta media total es cero) sólo si I de estos nuevos parámetros están determinados de manera independiente, así que existen muchos parámetros independientes como los hubo antes. En función de μ y las α_i, el modelo se vuelve

$$X_{ij} = \mu + \alpha_1 + \epsilon_{ij} \quad (i = 1, \ldots, I; \quad j = 1, \ldots, J)$$

En el capítulo 11 se desarrollarán modelos análogos para ANOVA multifactorial. La afirmación de que las μ_i son idénticas es equivalente a la igualdad de las α_i y como $\sum\alpha_i = 0$, la hipótesis nula se vuelve

$$H_0\colon \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0$$

Recuerde que MSTr es un estimador insesgado de σ^2 cuando H_0 es verdadera aunque de lo contrario tiende a sobrestimar σ^2. Más precisamente:

$$E(\text{MSTr}) = \sigma^2 + \frac{J}{I-1}\sum\alpha_i^2$$

Cuando H_0 es verdadera, $\sum\alpha_i^2 = 0$ de modo que $E(\text{MSTr}) = \sigma^2$ (MSE es insesgada sea o no verdadera H_0). Si $\sum\alpha_i^2$ se utiliza como medida del grado al cual H_0 es falsa, entonces un valor más grande de $\sum\alpha_i^2$ provocará una mayor tendencia de que MSTr sobrestime σ^2. En el siguiente capítulo, se utilizarán fórmulas para cuadrados de las medias esperadas en modelos multifactoriales a fin de sugerir cómo formar proporciones F para probar varias hipótesis.

Comprobación de la fórmula para *E*(MSTr) Con cualquier variable aleatoria, $E(Y^2) = V(Y) + [E(Y)]^2$, por lo tanto

$$\begin{aligned}
E(\text{SSTr}) &= E\left(\frac{1}{J}\sum_i X_{i\cdot}^2 - \frac{1}{IJ}X_{\cdot\cdot}^2\right) = \frac{1}{J}\sum_i E(X_{i\cdot}^2) - \frac{1}{IJ}E(X_{\cdot\cdot}^2)\\
&= \frac{1}{J}\sum_i\{V(X_{i\cdot}) + [E(X_{i\cdot})^2]\} - \frac{1}{IJ}\{V(X_{\cdot\cdot}) + [E(X_{\cdot\cdot})]^2\}\\
&= \frac{1}{J}\sum_i\{J\sigma^2 + [J(\mu+\alpha_i)]^2\} - \frac{1}{IJ}[IJ\sigma^2 + (IJ\mu)^2]\\
&= I\sigma^2 + IJ\mu^2 + 2\mu J\sum_i\alpha_i + J\sum_i\alpha_i^2 - \sigma^2 - IJ\mu^2\\
&= (I-1)\sigma^2 + J\sum_i\alpha_i^2 \quad (\text{puesto que}\sum\alpha_i = 0)
\end{aligned}$$

El resultado se deriva entonces de la relación $\text{MSTr} = \text{SSTr}/(I-1)$. ■

β para la prueba F

Considérese un conjunto de valores de parámetro $\alpha_1, \alpha_2, \ldots, \alpha_I$ con los cuales H_0 no es verdadera. La probabilidad de un error de tipo II, β, es la probabilidad de que H_0 no sea rechazada cuando ese conjunto es el conjunto de valores verdaderos. Se podría pensar que β tendría que ser determinada por separado para cada configuración diferente de α_i. Afortunadamente, como β para la prueba F depende de las α_i y σ^2 sólo mediante $\sum\alpha_i^2/\sigma^2$, se puede evaluar al mismo tiempo para muchas alternativas diferentes. Por ejemplo, $\sum\alpha_i^2 = 4$ para cada uno de los siguientes conjuntos de α_i con los cuales H_0 es falsa, así que β es idéntica para las tres alternativas:

1. $\alpha_1 = -1, \alpha_2 = -1, \alpha_3 = 1, \alpha_4 = 1$
2. $\alpha_1 = -\sqrt{2}, \alpha_2 = \sqrt{2}, \alpha_3 = 0, \alpha_4 = 0$
3. $\alpha_1 = -\sqrt{3}, \alpha_2 = \sqrt{1/3}, \alpha_3 = \sqrt{1/3}, \alpha_4 = \sqrt{1/3}$

La cantidad $J\sum\alpha_i^2/\sigma^2$ se llama **parámetro de no centralidad** para ANOVA unidireccional (debido a que cuando H_0 es falsa el estadístico de prueba tiene una distribución *F no centralizada* con éste como uno de sus parámetros) y β es una función decreciente del valor de este parámetro. Por lo tanto, con valores fijos de σ^2 y J, es más probable que la hipótesis nula sea rechazada para alternativas alejadas de H_0 ($\sum\alpha_i^2$ grande) que para alternativas próximas a H_0. Con un valor fijo de $\sum\alpha_i^2$, β se reduce a medida que el tamaño de muestra J en cada tratamiento se incrementa y aumenta a medida que la varianza σ^2 se

incrementa (puesto que una variabilidad subyacente más grande dificulta detectar cualquier alejamiento dado con respecto a H_0).

Como el cálculo a mano de β y la determinación del tamaño de muestra para la prueba F son bastante difíciles (como en el caso de pruebas t), los estadísticos han construido conjuntos de curvas donde se puede obtener β. En las figuras 10.5* y 10.6* se muestran conjuntos de curvas para $\nu_1 = 3$ y $\nu_1 = 4$ grados de libertad asociados con el numerador, respectivamente. Una vez que se especifican los valores de σ^2 y las α_i para los cuales se desea β, éstos se utilizan para calcular el valor de ϕ, donde $\phi^2 = (J/I)\sum\alpha_i^2/\sigma^2$. Luego se localiza el valor de ϕ en el conjunto de curvas apropiado sobre el eje horizontal, se sube hasta la curva asociada con los ν_2 grados de libertad asociados con el error y se localiza el valor de la potencia sobre el eje vertical. Finalmente, $\beta = 1 -$ potencia.

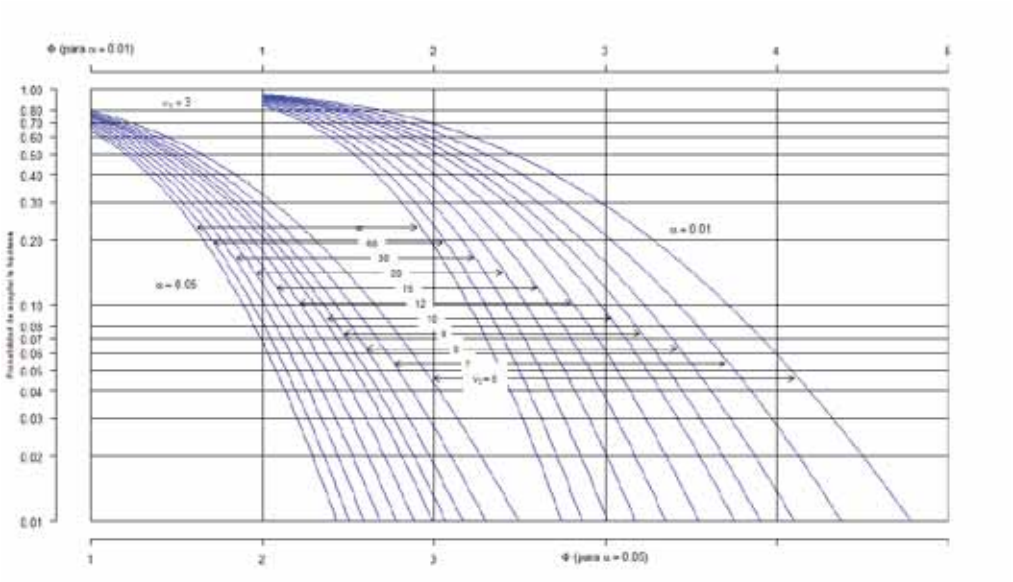


Figura 10.5 Curvas de probabilidad β, (potencia 1-β) para la prueba ANOVA F $\nu_1 = 3$

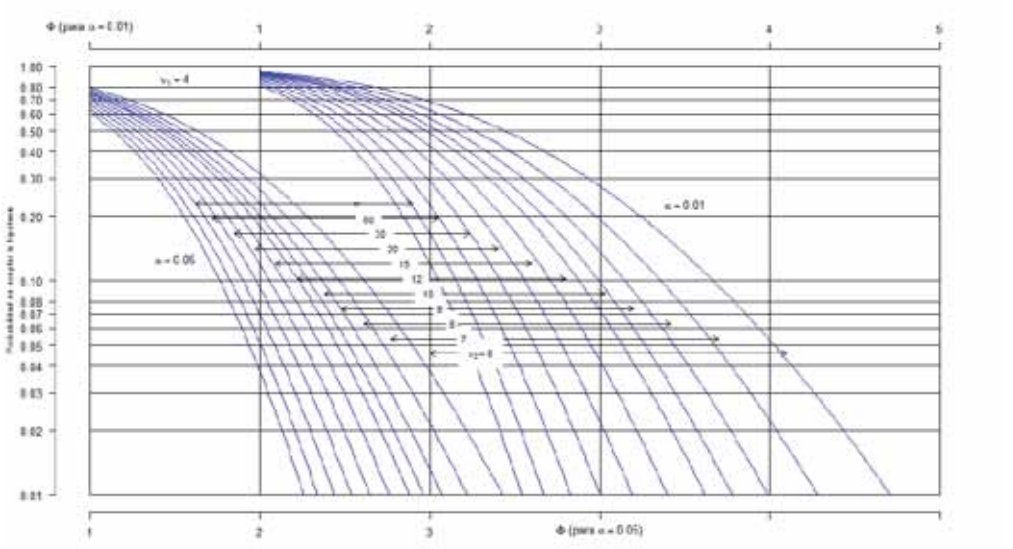


Figura 10.6 Curvas de probabilidad β, (potencia 1-β) para la prueba ANOVA F $\nu_1 = 4$

Ejemplo 10.8 Se tienen que investigar los efectos de cuatro tratamientos térmicos diferentes en el punto de cedencia (tons/pulg^2) de lingotes de acero. Un total de ocho lingotes se fundirán utilizando cada tratamiento. Suponga que la desviación estándar verdadera del punto de cedencia con cualquiera de los cuatro tratamientos es $\sigma = 1$. ¿Qué tan probable es que H_0 no será rechazada a un nivel de .05 si tres de los tratamientos tienen el mismo punto de cedencia esperado y el otro tiene un punto de cedencia esperado que es 1 ton/pulg^2 más grande que el valor común de los otros tres (es decir, la cuarta cedencia está en promedio a 1 desviación estándar por encima de aquellas con los primeros tres tratamientos)?

Suponga que $\mu_1 = \mu_2 = \mu_3$ y $\mu_4 = \mu_1 + 1$, $\mu = (\sum \mu_i)/4 = \mu_1 + \frac{1}{4}$. Entonces $\alpha_1 = \mu_1 - \mu = -\frac{1}{4}$, $\alpha_2 = -\frac{1}{4}$, $\alpha_3 = -\frac{1}{4}$, $\alpha_4 = \frac{3}{4}$ por lo tanto

$$\phi^2 = \frac{8}{4}\left[\left(-\frac{1}{4}\right)^2 + \left(-\frac{1}{4}\right)^2 + \left(-\frac{1}{4}\right)^2 + \left(\frac{3}{4}\right)^2\right] = \frac{3}{2}$$

y $\phi = 1.22$. Los grados de libertad para la prueba F son $\nu_1 = I - 1 = 3$ y $\nu_2 = I(J - 1) = 28$, si se interpola visualmente entre $\nu_2 = 20$ y $\nu_2 = 30$ se obtiene una potencia $\approx .47$ y $\beta \approx .53$. Esta β es algo grande, así que se podría incrementar el valor de J. ¿Cuántos lingotes de cada tipo se requerirían para dar $\beta \approx .05$ para la alternativa considerada? Probando diferentes valores de J, se puede verificar que $J = 24$ satisfará el requerimiento, pero cualquier J más pequeño no lo hará. ■

Como una alternativa del uso de curvas de potencia, el programa estadístico SAS incluye una función que calcula el área acumulada bajo una curva F no centralizada (se ingresa F_α, grados de libertad asociados con el numerador, grados de libertad asociados con el denominador y ϕ^2) y esta área es β. Minitab hace esto y también algo un tanto diferente. Se le pide al usuario que especifique la diferencia máxima entre las μ_i y no entre las medias individuales. Por ejemplo, se podría desear calcular la potencia de la prueba cuando $I = 4$, $\mu_1 = 100$, $\mu_2 = 101$, $\mu_3 = 102$ y $\mu_4 = 106$. Entonces la diferencia máxima es $106 - 100 = 6$. Sin embargo, la potencia no sólo depende de esta diferencia máxima sino de los valores de todas las μ_i. En esta situación Minitab calcula el valor de potencia más pequeño posible sujeto a $\mu_1 = 100$ y $\mu_4 = 106$, lo cual ocurre cuando las otras dos μ se encuentran a la mitad entre 100 y 106. Si esta potencia es de .85, entonces se puede decir que la potencia es de por lo menos .85 y β es cuando mucho de .15 cuando las dos μ más extremas están separadas por 6 (el tamaño de muestra común, α y σ también deben ser especificados). El programa determinará también el tamaño de muestra común requerido si la diferencia máxima y la potencia mínima están especificadas.

Relación de la prueba *F* con la prueba *t*

Cuando el número de tratamientos o poblaciones es $I = 2$, todas las fórmulas y resultados conectados con la prueba F siguen teniendo sentido, así que se puede utilizar ANOVA para probar H_0: $\mu_1 = \mu_2$ contra H_a: $\mu_1 \neq \mu_2$. En este caso, también se puede utilizar una prueba t con dos muestras de dos colas. En la sección 9.3, se mencionó la prueba t agrupada, la cual requiere varianzas iguales, como alternativa del procedimiento t con dos muestras. Se puede demostrar que la prueba F ANOVA unifactorial y la prueba t agrupada de dos colas son equivalentes; con cualquier conjunto de datos dado, los valores P para las dos pruebas serán idénticos, así que se llegará a la misma conclusión con cualquier prueba.

La prueba t con dos muestras es más flexible que la prueba F cuando $I = 2$ por dos razones. En primer lugar, es válida sin la suposición de que $\sigma_1 = \sigma_2$; en segundo lugar, puede ser utilizada para probar H_a: $\mu_1 > \mu_2$ (una prueba t de cola superior) o H_a: $\mu_1 < \mu_2$ así como también H_a: $\mu_1 \neq \mu_2$. En el caso de $I \geq 3$, desafortunadamente no existe un procedimiento de prueba general que tenga buenas propiedades sin la suposición de varianzas iguales.

Tamaños de muestra desiguales

Cuando los tamaños de muestra de cada población o tratamiento no son iguales, sean $J_1, J_2, \ldots, J_I$ los tamaños de muestra I y sea $n = \Sigma_i J_i$ el número total de observaciones. El recuadro adjunto da fórmulas ANOVA y el procedimiento de prueba.

$$\text{SST} = \sum_{i=1}^{I}\sum_{j=1}^{J_i}(X_{ij} - \bar{X}_{\cdot\cdot})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J_i}X_{ij}^2 - \frac{1}{n}X_{\cdot\cdot}^2 \quad \text{gl} = n - 1$$

$$\text{SSTr} = \sum_{i=1}^{I}\sum_{j=1}^{J_i}(\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2 = \sum_{i=1}^{I}\frac{1}{J_i}X_{i\cdot}^2 - \frac{1}{n}X_{\cdot\cdot}^2 \quad \text{gl} = I - 1$$

$$\text{SSE} = \sum_{i=1}^{I}\sum_{j=1}^{J_i}(X_{ij} - \bar{X}_{i\cdot})^2 = \text{SST} - \text{SSTr} \quad \text{gl} = \sum(J_i - 1) = n - I$$

Valor estadístico de prueba:

$$f = \frac{\text{MSTr}}{\text{MSE}} \quad \text{donde MSTr} = \frac{\text{SSTr}}{I - 1} \quad \text{MSE} = \frac{\text{SSE}}{n - I}$$

Región de rechazo: $f \geq F_{\alpha,I-1,n-I}$

Ejemplo 10.9 El artículo "On the Development of a New Approach for the Determination of Yield Strength in Mg-based Alloys" (*Light Metal Age*, octubre de 1998: 51–53) presentó los siguientes datos sobre módulo elástico (GPa) obtenidos por medio de un nuevo método ultrasónico con especímenes de cierta aleación producida mediante tres procesos de fundición diferentes.

									J_i	$x_{i\cdot}$	$\bar{x}_{i\cdot}$
Moldeado permanente	45.5	45.3	45.4	44.4	44.6	43.9	44.6	44.0	8	357.7	44.71
Fundición a troquel	44.2	43.9	44.7	44.2	44.0	43.8	44.6	43.1	8	352.5	44.06
Moldeado en yeso	46.0	45.9	44.8	46.2	45.1	45.5			6	273.5	45.58
									22	983.7	

Sean μ_1, μ_2 y μ_3 los módulos elásticos promedio verdaderos con los tres procesos diferentes en las circunstancias dadas. Las hipótesis pertinentes son H_0: $\mu_1 = \mu_2 = \mu_3$ contra H_a: por lo menos dos de las μ_i son diferentes. El estadístico de prueba es, desde luego, $F = \text{MSTr/MSE}$, basado en $I - 1 = 2$ grados de libertad asociados con el numerador y $n - I = 22 - 3 = 19$ grados de libertad asociados con el denominador. Las cantidades pertinentes incluyen

$$\sum\sum x_{ij}^2 = 43{,}998.73 \quad \text{FC} = \frac{983.7^2}{22} = 43{,}984.80$$

$$\text{SST} = 43{,}998.73 - 43{,}984.80 = 13.93$$

$$\text{SSTr} = \frac{357.7^2}{8} + \frac{352.5^2}{8} + \frac{273.5^2}{6} - 43{,}984.84 = 7.93$$

$$\text{SSE} = 13.93 - 7.93 = 6.00$$

Los cálculos restantes se muestran en la tabla ANOVA adjunta. Como $F_{.001,2,19} = 10.16 < 12.56 = f$, el valor P es más pequeño que .001. Por consiguiente, la hipótesis nula deberá ser rechazada a cualquier nivel de significancia razonable; existe evidencia convincente

para concluir que el módulo elástico promedio verdadero en cierta forma depende de qué proceso de fundición se utilice.

Origen de la variación	Grados de libertad	Suma de los cuadrados	Cuadrado de la media	f
Tratamientos	2	7.93	3.965	12.56
Error	19	6.00	.3158	
Total	21	13.93		

Existe más controversia entre estadísticos con respecto a qué procedimiento de comparaciones múltiples utilizar cuando los tamaños de muestra son desiguales que los que existen en el caso de tamaños de muestra iguales. El procedimiento que aquí se presenta lo recomienda el excelente libro *Beyond ANOVA*: *Basics of Applied Statistics* (véase la bibliografía del capítulo) para usarse cuando los I tamaños de muestra $J_1, J_2, \ldots J_I$ están razonablemente cerca uno de otro ("desequilibrio leve"). Modifica el método de Tukey por medio de promedios de pares $1/J_i$ en lugar de $1/J$.

Sea

$$w_{ij} = Q_{\alpha,I,n-I} \cdot \sqrt{\frac{\text{MSE}}{2}\left(\frac{1}{J_i} + \frac{1}{J_j}\right)}$$

Entonces la probabilidad es *aproximadamente* $1 - \alpha$ de que

$$\bar{X}_{i\cdot} - \bar{X}_{j\cdot} - w_{ij} \le \mu_i - \mu_j \le \bar{X}_{i\cdot} - \bar{X}_{j\cdot} + w_{ij}$$

con cada i y j ($i = 1, \ldots, I$ y $j = 1, \ldots, I$) con $i \ne j$.

El nivel de confianza simultáneo de $100(1 - \alpha)\%$ es sólo aproximado y no exacto ya que se determinó con tamaños de muestra iguales. El método de subrayado puede seguir siendo utilizado, pero ahora el factor w_{ij} utilizado para decidir si $\bar{x}_{i\cdot}$ y $\bar{x}_{j\cdot}$ pueden ser conectadas dependerá de J_i y J_j.

Ejemplo 10.10 (Continuación del ejemplo 10.9)

Los tamaños de muestra para los datos de módulo elástico fueron $J_1 = 8$, $J_2 = 8$, $J_3 = 6$ e $I = 3$, $n - I = 19$, MSE = .316. Un nivel de confianza simultáneo aproximadamente de 95% requiere $Q_{.05,3,19} = 3.59$, de donde

$$w_{12} = 3.59\sqrt{\frac{.316}{2}\left(\frac{1}{8} + \frac{1}{8}\right)} = .713, \quad w_{13} = .771 \quad w_{23} = .771$$

Como $\bar{x}_{1\cdot} - \bar{x}_{2\cdot} = 44.71 - 44.06 = .65 < w_{12}$, μ_1 y μ_2 se consideran no significativamente diferentes. El esquema de subrayado adjunto muestra que en apariencia μ_1 y μ_3 difieren de manera significativa, como lo hacen μ_2 y μ_3.

2. Troquel	1. Permanente	3. Yeso
44.06	44.71	45.58

Transformación de datos

El uso de métodos ANOVA puede ser invalidado por diferencias sustanciales en las varianzas $\sigma_1^2, \ldots, \sigma_I^2$ las que hasta ahora han sido supuestas iguales con valor común de σ^2). En ocasiones sucede que $V(X_{ij}) = \sigma_i^2 = g(\mu_i)$, una función conocida de μ_i (de modo que cuando H_0 es falsa, las varianzas no son iguales). Por ejemplo, si X_{ij} tiene una distribución

de Poisson con parámetro λ_i (aproximadamente normal si $\lambda_i \geq 10$), entonces $\mu_i = \lambda_i$ y $\sigma_i^2 = \lambda_i$, de modo que $g(\mu_i) = \mu_i$ es la función conocida. En tales casos, a menudo se pueden transformar las X_{ij} en $h(X_{ij})$ de modo que tengan varianzas iguales de manera aproximada (al mismo tiempo que las variables transformadas permanecen aproximadamente normales) y luego se puede utilizar la prueba F con las observaciones transformadas. La idea clave al seleccionar $h(\cdot)$ es que con frecuencia $V[h(X_{ij})] \approx V(X_{ij}) \cdot [h'(\mu_i)]^2 = g(\mu_i) \cdot [h'(\mu_i)]^2$. Se desea determinar la función $h(\cdot)$ con la cual $g(\mu_i) \cdot [h'(\mu_i)]^2 = c$ (una constante) con cada i.

PROPOSICIÓN

Si $V(X_{ij}) = g(\mu_i)$, una función conocida de μ_i, entonces una transformación $h(X_{ij})$ que "estabilice la varianza" de modo que $V[h(X_{ij})]$ sea aproximadamente la misma con cada i está dada por $h(x) \propto \int [g(x)]^{-1/2}\, dx$.

En el caso Poisson, $g(x) = x$, de modo que $h(x)$ deberá ser proporcional a $\int x^{-1/2}\, dx = 2x^{1/2}$. Así pues los datos Poisson deberán ser cambiados a $h(x_{ij}) = \sqrt{x_{ij}}$ antes del análisis.

Un modelo de efectos aleatorios

Se ha supuesto que los problemas unifactoriales considerados hasta ahora son ejemplos de un modelo ANOVA de **efectos fijos.** Con esto se quiere decir que los niveles elegidos del factor en estudio son los únicos considerados pertinentes por el experimentador. El modelo de efectos fijos unifactorial es

$$X_{ij} = \mu + \alpha_i + \epsilon_{ij} \qquad \sum \alpha_i = 0 \tag{10.6}$$

donde las ϵ_{ij} son aleatorias y tanto μ como las α_i son parámetros fijos.

En algunos problemas unifactoriales, los niveles particulares estudiados por el experimentador se seleccionan, mediante diseño o mediante muestreo, de una gran población de niveles. Por ejemplo, para estudiar los efectos en tiempo de desempeño de una tarea por la utilización de diferentes operarios en una máquina particular, se podría seleccionar una muestra de cinco operarios de un gran conjunto de operarios. Asimismo, se podría estudiar el efecto del pH del suelo en la cosecha de plantas de maíz utilizando suelos con cuatro valores de pH específicos seleccionados de entre los muchos niveles de pH posibles. Cuando los niveles utilizados se seleccionan al azar de entre una gran población de niveles posibles, se dice que el factor es aleatorio y no fijo, y el modelo de efectos fijos (10.6) ya no es apropiado. Un modelo de **efectos aleatorios análogos** se obtiene reemplazando las α_i fijas en (10.6) por variables aleatorias.

$$X_{ij} = \mu + A_i + \epsilon_{ij} \quad \text{con } E(A_i) = E(\epsilon_{ij}) = 0$$

$$V(\epsilon_{ij}) = \sigma^2 \qquad V(A_i) = \sigma_A^2 \tag{10.7}$$

todas las A_i y ϵ_{ij} normalmente distribuidas e independientes una de otra.

La condición $E(A_i) = 0$ en (10.7) es similar a la condición $\Sigma \alpha_i = 0$ en (10.6); manifiesta que el efecto esperado o promedio del $i$ésimo nivel medido como un alejamiento de μ es cero.

Para el modelo de efectos aleatorios (10.7), la hipótesis de ningunos efectos debido a los diferentes niveles es H_0: $\sigma_A^2 = 0$, la cual expresa que los diferentes niveles del factor con-

tribuyen con nada a la variabilidad de la respuesta. *Aunque las hipótesis en los modelos de efectos fijos unifactoriales y aleatorios son diferentes, se prueban en exactamente la misma manera*, formando $F = \text{MSTr/MSE}$ y rechazando H_0, si $f \geq F_{\alpha, I-1, n-I}$. Esto se justifica de manera intuitiva al observar que $E(\text{MSE}) = \sigma^2$ (como para efectos fijos), mientras que

$$E(\text{MSTr}) = \sigma^2 + \frac{1}{I-1}\left(n - \frac{\sum J_i^2}{n}\right)\sigma_A^2 \tag{10.8}$$

donde $J_1, J_2, \ldots, J_I$ son los tamaños de muestra y $n = \sum J_i$. El factor entre paréntesis en el lado derecho de (10.8) es no negativo, de modo que de nuevo $E(\text{MSTr}) = \sigma^2$ si H_0 es verdadera y $E(\text{MSTr}) > \sigma^2$ si H_0 es falsa.

Ejemplo 10.11 El estudio de fuerzas y esfuerzos no destructivos en materiales aporta información importante para el diseño de ingeniería eficiente. El artículo "Zero-Force Travel-Time Parameters for Ultrasonic Head-Waves in Railroad Rail" (*Materials Evaluation*, 1985: 854–858) reporta sobre un estudio de tiempo de recorrido de cierto tipo de onda que produce el esfuerzo longitudinal de rieles utilizados en vías de ferrocarril. Se realizaron tres mediciones en cada uno de seis rieles seleccionados al azar de una población de rieles. Los investigadores utilizaron ANOVA de efectos aleatorios para decidir si algo de la variación del tiempo de recorrido podía ser atribuido a la "variabilidad entre rieles". Los datos se dan en la tabla adjunta (cada valor, en nanosegundos, se obtuvo de restar 36.1 μ de la observación original) junto con la tabla ANOVA derivada. El valor de la proporción f es altamente significativo, así que H_0: $\sigma_A^2 = 0$ es rechazada a favor de la conclusión de que las diferencias entre rieles provocan la variabilidad del tiempo de recorrido.

				$x_{i\cdot}$
1:	55	53	54	162
2:	26	37	32	95
3:	78	91	85	254
4:	92	100	96	288
5:	49	51	50	150
6:	80	85	83	248
			$x_{\cdot\cdot}$ =	1197

Origen de la variación	Grados de libertad	Suma de los cuadrados	Cuadrado de la media	f
Tratamientos	5	9310.5	1862.1	115.2
Error	12	194.0	16.17	
Total	17	9504.5		

EJERCICIOS Sección 10.3 (22–34)

22. Los datos siguientes se refieren a la cosecha de tomates (kg/parcela) con cuatro niveles de salinidad diferentes; el *nivel de salinidad* aquí se refiere a la conductividad eléctrica (CE), donde los niveles seleccionados fueron CE = 1.6, 3.8, 6.0 y 10.2 nmhos/cm.

1.6:	59.5	53.3	56.8	63.1	58.7
3.8:	55.2	59.1	52.8	54.5	
6.0:	51.7	48.8	53.9	49.0	
10.2:	44.6	48.5	41.0	47.3	46.1

Use la prueba F al nivel $\alpha = .05$ para probar en cuanto a cualquier diferencia en la cosecha promedio verdadera debido a los distintos niveles de salinidad.

23. Aplique el método de Tukey modificado a los datos del ejercicio 22 para identificar diferencias significativas entre las μ_i.

24. Los datos de la tabla que resumen la actividad del músculo esquelético CS (nmol/min/mg) aparecieron en el artículo "Impact of Lifelong Sedentary Behavior on Mitochondrial Function of Mice Skeletal Muscle" (*J. of Gerontology*, 2009: 927–939):

	Jóvenes	Adultos sedentarios	Adultos activos
Tamaño muestral	10	8	10
Media muestral	46.68	47.71	58.24
DE muestral	7.16	5.59	8.43

Lleve a cabo una prueba para decidir si la actividad promedio real es diferente para los tres grupos. En su caso, analice las diferencias entre las medias con un método de comparaciones múltiples.

25. Los lípidos aportan mucha de la energía dietética en los cuerpos de bebés y niños. Existe un interés creciente en la calidad del suministro de lípido dietético durante la infancia como un importante determinante del crecimiento, desarrollo visual y nervioso y salud a largo plazo. El artículo "Essential Fat Requirements of Preterm Infants" (*Amer. J. of Clinical Nutrition*, 2000: 245S–250S) reportó los siguientes datos sobre grasas poliinsaturadas (%) para bebés que fueron asignados al azar a cuatro regímenes de alimentación diferentes: leche materna, fórmula basada en aceite de maíz (CO), fórmula basada en aceite de soya (SO) o fórmula basada en aceite de soya y marino (SMO).

Régimen	Tamaño de muestra	Media muestral	Desv. estándar muestral
Leche materna	8	43.0	1.5
CO	13	42.4	1.3
SO	17	43.1	1.2
SMO	14	43.5	1.2

a. ¿Qué suposiciones se deben hacer sobre las cuatro distribuciones de grasa poliinsaturada antes de realizar un ANOVA unifactorial para decidir si existen diferencias en el contenido de grasa promedio verdadero?
b. Realice la prueba sugerida en el inciso (a). ¿Qué se puede decir sobre el valor *P*?

26. Se analizaron muestras de seis marcas diferentes de margarina dietética/imitación para determinar el nivel de ácidos grasos poliinsaturados fisiológicamente activos (PARFUA, por sus siglas en inglés, en porcentajes) y se obtuvieron los siguientes resultados:

Imperial	14.1	13.6	14.4	14.3	
Parkay	12.8	12.5	13.4	13.0	12.3
Blue Bonnet	13.5	13.4	14.1	14.3	
Chiffon	13.2	12.7	12.6	13.9	
Mazola	16.8	17.2	16.4	17.3	18.0
Fleischmann's	18.1	17.2	18.7	18.4	

(Los números precedentes son ficticios, aunque las medias muestrales concuerdan con los datos reportados en el ejemplar de enero de 1975 de *Consumer Reports*.)
a. Use ANOVA para probar las diferencias entre los porcentajes de ácidos grasos poliinsaturados fisiológicamente activos para las distintas marcas.
b. Calcule intervalos de confianza para todas las $(\mu_i - \mu_j)$.
c. Mazola y Fleischmann's son aceites de maíz, en tanto que los demás son de soya. Calcule un intervalo de confianza para

$$\frac{(\mu_1 + \mu_2 + \mu_3 + \mu_4)}{4} - \frac{(\mu_5 + \mu_6)}{2}$$

[*Sugerencia*: modifique la expresión para $V(\hat{\theta})$ que condujo a (10.5) en la sección previa.]

27. Aunque el té es la bebida que más se consume en el mundo después del agua, se sabe poco sobre su valor nutricional. La folacina es la única vitamina B presente en cualquier cantidad significativa de té y avances recientes en métodos de ensayo han determinado con precisión el contenido de folacina factible. Considere los datos adjuntos sobre contenido de folacina en especímenes seleccionados al azar de las cuatro marcas líderes de té verde.

1:	7.9	6.2	6.6	8.6	8.9	10.1	9.6
2:	5.7	7.5	9.8	6.1	8.4		
3:	6.8	7.5	5.0	7.4	5.3	6.1	
4:	6.4	7.1	7.9	4.5	5.0	4.0	

(Los datos están basados en "Folacin Content of Tea", *J. of the Amer. Dietetic Assoc*., 1983: 627–632.) ¿Sugieren estos datos que el contenido de folacina promedio verdadero es el mismo para todas las marcas?
a. Realice una prueba con $\alpha = .05$ con el método del valor *P*.
b. Evalúe la factibilidad de cualquier suposición requerida para su análisis en el inciso (a).
c. Realice un análisis de comparaciones múltiples para identificar diferencias significativas entre marcas.

28. Para un ANOVA unifactorial con tamaños de muestra J_i ($i = 1, 2, \ldots I$) demuestre que $\text{SSTr} = \sum J_i(\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2 = \sum J_i \bar{X}_{i\cdot}^2 - n\bar{X}_{\cdot\cdot}^2$, donde $n = \Sigma J_i$.

29. Cuando los tamaños de muestra son iguales ($J_i = J$), los parámetros $\alpha_1, \alpha_2, \ldots \alpha_I$ de la parametrización alternativa están restringidos por $\Sigma\alpha_i = 0$. Con tamaños de muestra desiguales, la restricción más natural es $\Sigma J_i\alpha_i = 0$. Use esto para demostrar que

$$E(\text{MSTr}) = \sigma^2 + \frac{1}{I-1}\sum J_i\alpha_i^2$$

¿Cuál es E(MSTr) cuando H_0 es verdadera? [Esta expectativa es correcta si $\sum J_i\alpha_i = 0$ es reemplazada por la restricción $\sum\alpha_i = 0$ (o por cualquier otra restricción lineal sobre las α_i utilizadas para reducir el modelo a I parámetros independientes), pero $\sum J_i\alpha_i = 0$ simplifica el álgebra y produce estimaciones naturales para los parámetros del modelo (en particular $\hat{\alpha}_i = \bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot}$).]

30. Reconsidere el ejemplo 10.8 que implica una investigación de los efectos de diferentes tratamientos térmicos sobre el punto de cedencia de lingotes de acero.
a. Si $J = 8$ y $\sigma = 1$, ¿cuál es β para una prueba F a un nivel de .05 cuando $\mu_1 = \mu_2, \mu_3 = \mu_1 - 1$ y $\mu_4 = \mu_1 + 1$?
b. Con la alternativa del inciso (a), ¿qué valor de J es necesario para obtener $\beta = .05$?
c. Si existen $I = 5$ tratamientos térmicos, $J = 10$ y $\sigma = 1$, ¿cuál es β para la prueba F a un nivel de .05 cuando cuatro de las μ_i son iguales y la quinta difiere en 1 de las otras cuatro?

31. Cuando los tamaños de muestra no son iguales, el parámetro de no centralidad es $\sum J_i\alpha_i^2/\sigma^2$ y $\phi^2 = (1/I)\sum J_i\alpha_i^2/\sigma^2$. Remitiéndose al ejercicio 22, ¿cuál es la potencia de la prueba cuando $\mu_2 = \mu_3, \mu_1 = \mu_2 - \sigma$ y $\mu_4 = \mu_2 + \sigma$?

32. En un experimento para comparar la calidad de cuatro marcas diferentes de cinta para grabar de carrete a carrete, se seleccionaron cuatro carretes de 2400 pies de cada marca (A–D) y se determinó el número de imperfecciones en cada uno.

A:	10	5	12	14	8
B:	14	12	17	9	8
C:	13	18	10	15	18
D:	17	16	12	22	14

Se cree que el número de imperfecciones tiene aproximadamente una distribución Poisson para cada marca. Analice los datos al nivel .01 con objeto de ver si el número esperado de imperfecciones por carrete es el mismo para cada marca.

33. Suponga que X_{ij} es una variable binomial con parámetros n y p_i (así que es aproximadamente normal cuando $np_i \geq 10$ y $nq_i \geq 10$. Entonces como $\mu_i = np_i$, $V(X_{ij}) = \sigma_i^2 = np_i(1 - p_i) = \mu_i(1 - \mu_i/n)$. ¿Cómo se deberán transformar las X_{ij} de modo que se estabilice la varianza? [*Sugerencia*: $g(\mu_i) = \mu_i(1 - \mu_i/n)$.]

34. Simplifique $E(\text{MSTr})$ para el modelo de efectos aleatorios cuando $J_1 = J_2 = \cdots = J_I = J$.

EJERCICIOS SUPLEMENTARIOS (35–46)

35. Se realizó un experimento para comparar las velocidades de flujo de cuatro tipos diferentes de boquilla.

a. Los tamaños de muestra fueron 5, 6, 7 y 6, respectivamente y los cálculos dieron $f = 3.68$. Formule y pruebe las hipótesis pertinentes con $\alpha = .01$.

b. El análisis de los datos con un programa estadístico dio el valor $P = .029$. Al nivel .01, ¿qué concluiría y por qué?

36. El artículo "Computer-Assisted Instruction Augmented with Planned Teacher/Student Contacts" (*J. of the Exp. Educ.*, Invierno, 1980–1981: 120–126) comparó cinco métodos diferentes de enseñar estadística descriptiva. Los cinco fueron discusión y conferencia tradicionales (L/D), instrucción con libro de texto programado (R), texto programado con conferencias (R/L), instrucción con computadora (C) e instrucción con computadora y conferencias (C/L). Cuarenta y cinco estudiantes fueron asignados al azar, 9 a cada método. Después de completar el curso, los estudiantes resolvieron un examen de 1 hora. Además, se administró una prueba de retención de 10 minutos 6 semanas después. Los resultados son los siguientes.

	Examen		Prueba de retención	
Método	$\bar{x}_{i\cdot}$	s_i	$\bar{x}_{i\cdot}$	s_i
L/D	29.3	4.99	30.20	3.82
R	28.0	5.33	28.80	5.26
R/L	30.2	3.33	26.20	4.66
C	32.4	2.94	31.10	4.91
C/L	34.2	2.74	30.20	3.53

La media grande del examen fue 30.82 y la de la prueba de retención fue 29.30.

a. ¿Sugieren estos datos que existe diferencia entre los cinco métodos de enseñanza con respecto a la calificación del examen media verdadera? Use $\alpha = .05$.

b. Con un nivel de significancia de .05, pruebe la hipótesis nula de ninguna diferencia entre las calificaciones de la prueba de retención media verdadera para los cinco métodos de enseñanza distintos.

37. Numerosos factores contribuyen al funcionamiento suave de un motor eléctrico ("Increasing Market Share Through Improved Product and Process Design: An Experimental Approach", *Quality Engineering*, 1991: 361–369). En particular, es deseable mantener el ruido del motor y vibraciones a un mínimo. Para estudiar el efecto que la marca de los cojinetes tiene en la vibración del motor, se examinaron cinco marcas diferentes de cojinetes instalando cada tipo de cojinete en muestras aleatorias distintas de seis motores. Se registró la cantidad de vibración del motor (medida en micrones) cuando cada uno de los 30 motores estaba funcionando. Los datos de este estudio se dan a continuación. Formule y pruebe las hipótesis pertinentes a un nivel de significancia de .05 y luego realice un análisis de comparaciones múltiples si es apropiado.

							Media
1:	13.1	15.0	14.0	14.4	14.0	11.6	13.68
2:	16.3	15.7	17.2	14.9	14.4	17.2	15.95
3:	13.7	13.9	12.4	13.8	14.9	13.3	13.67
4:	15.7	13.7	14.4	16.0	13.9	14.7	14.73
5:	13.5	13.4	13.2	12.7	13.4	12.3	13.08

38. Un artículo publicado en el diario científico británico *Nature* ("Sucrose Induction of Hepatic Hyperplasia in the Rat", 25 de agosto de 1972: 461) reporta sobre un experimento en el cual cada uno de cinco grupos compuestos de seis ratas fue puesto a dieta con un carbohidrato diferente. Al final del experimento, se determinó el contenido de ADN del hígado de cada rata (mg/g hígado), con los siguientes resultados:

Carbohidrato	$\bar{x}_{i\cdot}$
Almidón	2.58
Sucrosa	2.63
Fructosa	2.13
Glucosa	2.41
Maltosa	2.49

Suponiendo también que $\sum\sum x_{ij}^2 = 183.4$, ¿indican estos datos que el tipo de carbohidrato presente en la dieta afecta el contenido de ADN promedio verdadero? Construya una tabla ANOVA y use un nivel de significancia de .05.

39. Remitiéndose al ejercicio 38, construya un intervalo de confianza t para

$$\theta = \mu_1 - (\mu_2 + \mu_3 + \mu_4 + \mu_5)/4$$

que mide la diferencia entre el contenido de ADN promedio para la dieta de almidón y el promedio combinado para las otras cuatro dietas. ¿Incluye cero el intervalo resultante?

40. Remítase al ejercicio 38. ¿Cuál es β para la prueba cuando el contenido de ADN promedio verdadero es idéntico para las tres dietas y queda a 1 desviación estándar (σ) por debajo de este valor común para las otras dos dietas?

41. Se seleccionan al azar cuatro laboratorios (1–4) de una población grande y a cada uno se le pide que haga tres determina-

ciones del porcentaje de alcohol metílico en especímenes de un compuesto tomado de un solo lote. Basado en los datos adjuntos, ¿son las diferencias entre los laboratorios una causa de variación del porcentaje de alcohol metílico? Formule y pruebe las hipótesis pertinentes con un nivel de significancia de .05.

1:	85.06	85.25	84.87
2:	84.99	84.28	84.88
3:	84.48	84.72	85.10
4:	84.10	84.55	84.05

42. La frecuencia de parpadeo crítica (cff) es la frecuencia más alta (en ciclos/s) a la que una persona puede advertir el parpadeo en una fuente luminosa parpadeante. A frecuencias por encima de la frecuencia de parpadeo crítica, la fuente luminosa parece ser continua aun cuando en realidad parpadee. Una investigación realizada para ver si la frecuencia de parpadeo crítica promedio verdadera depende del color del iris arrojó los siguientes datos (con base en el artículo "The Effects of Iris Color in Critical Flicker Frequency", *J. of General Psych*., 1973: 91–95):

	Color del iris		
	1. Café	**2. Verde**	**3. Azul**
	26.8	26.4	25.7
	27.9	24.2	27.2
	23.7	28.0	29.9
	25.0	26.9	28.5
	26.3	29.1	29.4
	24.8		28.3
	25.7		
	24.5		
J_i	8	5	6
$x_{i\cdot}$	204.7	134.6	169.0
$\bar{x}_{i\cdot}$	25.59	26.92	28.17

$n = 19 \qquad x_{\cdot\cdot} = 508.3$

a. Formule y pruebe las hipótesis pertinentes a un nivel de significancia de .05 utilizando la tabla *F* para obtener un límite superior y/o inferior del valor *P*. [*Sugerencia*: $\sum\sum x_{ij}^2 = 13{,}659.67$ y FC = 13,598.36.]

b. Investigue las diferencias entre colores del iris con respecto a la frecuencia de parpadeo crítica media.

43. Sean $c_1, c_2, \ldots, c_I$ los números que satisfacen la expresión $\sum c_i = 0$. Entonces $\sum c_i \mu_i = c_1\mu_1 + \cdots + c_I\mu_I$ se llama *contraste* en las μ_i. Observe que con $c_1 = 1$, $c_2 = -1$, $c_3 = \cdots = c_I = 0$, $\sum c_i \mu_i = \mu_1 - \mu_2$, la cual implica que toda diferencia tomada por pares entre las μ_i es un contraste (también lo es, p. ej., $\mu_1 - .5\mu_2 - .5\mu_3$). Un método atribuido a Scheffé da intervalos de confianza simultáneos con nivel de confianza simultáneo de $100(1 - \alpha)\%$ para *todos* los contrastes posibles (¡un número infinito de ellos!) El intervalo para $\sum c_i\mu_i$ es

$$\sum c_i\bar{x}_{i\cdot} \pm (\sum c_i^2/J_i)^{1/2} \cdot [(I - 1) \cdot \text{MSE} \cdot F_{\alpha, I-1, n-I}]^{1/2}$$

Usando los datos del ejercicio 42, acerca de la frecuencia crítica de parpadeo, calcule los intervalos de Scheffé para los contrastes $\mu_1 - \mu_2$, $\mu_1 - \mu_3$, $\mu_2 - \mu_3$, y $.5\mu_1 + .5\mu_2 - \mu_3$ (este último contraste compara azul con el promedio de café y verde). ¿Cuál contraste parece diferir significativamente de 0 y por qué?

44. Cuatro tipos de morteros: mortero de cemento ordinario (OCM), mortero impregnado de polímero (PIM), mortero con resina (RM) y mortero con cemento y polímero (PCM), se sometieron a una prueba de compresión para medir resistencia (MPa). Tres observaciones de resistencia de cada tipo de mortero se dan en el artículo "Polymer Mortar Composite Matrices for Maintenance-Free Highly Durable Ferrocement" (*J. of Ferrocement*, 1984: 337–345) y se reproducen aquí. Construya una tabla ANOVA. Con un nivel de significancia de .05, determine si los datos sugieren que la resistencia media verdadera no es la misma para los cuatro tipos de mortero. Si determina que las resistencias medias verdaderas no son iguales, use el método de Tukey para identificar las diferencias significativas.

OCM	32.15	35.53	34.20
PIM	126.32	126.80	134.79
RM	117.91	115.02	114.58
PCM	29.09	30.87	29.80

45. Suponga que las x_{ij} están "codificadas" por $y_{ij} = cx_{ij} + d$. ¿Cómo se compara el valor del estadístico *F* calculado con las y_{ij} con el valor calculado con las x_{ij}? Justifique su aseveración.

46. En el ejemplo 10.11, reste $\bar{x}_{i\cdot}$ de cada observación en la *i*ésima muestra ($i = 1, \ldots, 6$) para obtener un conjunto de 18 residuos. Luego construya una gráfica de probabilidad normal y comente sobre la factibilidad de la suposición de normalidad.

Bibliografía

Miller, Rupert, *Beyond ANOVA*: *The Basics of Applied Statistics*, Wiley, Nueva York, 1986. Una excelente fuente de información sobre comprobación de suposiciones y métodos de análisis alternativos.

Montgomery, Douglas, *Design and Analysis of Experiments* (7a. ed.), Wiley, Nueva York, 2009. Una presentación muy al día de modelos y metodología ANOVA.

Neter, John, William Wasserman y Michael Kutner, *Applied Linear Statistical Models* (5a. ed.), Irwin, Homewood, IL., 2004. La segunda mitad de este libro contiene un estudio muy bien presentado de ANOVA; el nivel es comparable al del presente texto, aunque la discusión es más amplia, lo que hace del libro una excelente referencia.

Ott, R. Lyman y Michael Longnecker. *An Introduction to Statistical Methods and Data Analysis* (6a. ed.), Duxbury Press, Belmont, CA, 2010. Incluye varios capítulos sobre metodología ANOVA que puede ser leído con provecho por estudiantes que desean una exposición no muy matemática; incluye un capítulo muy bueno sobre varios métodos de comparaciones múltiples.

11 Análisis multifactorial de la varianza

INTRODUCCIÓN

En el capítulo previo se utilizó el análisis de la varianza (ANOVA) para probar en cuanto a igualdad de I medias de población diferentes o las respuestas promedio verdaderas asociadas con I niveles diferentes de un solo factor (alternativamente conocidos como I tratamientos diferentes). En muchas situaciones experimentales, existen dos o más factores que son de interés simultáneo. Este capítulo amplía los métodos del capítulo 10 para investigar tales situaciones multifactoriales.

En las dos primeras secciones, se concentra en el caso de dos factores. Se utilizará I para denotar el número de niveles del primer factor (A) y J para denotar el número de niveles del segundo factor (B). Entonces existen IJ combinaciones posibles compuestas de un nivel del factor A y uno del factor B. Cada una de tales combinaciones se conoce como tratamiento, de ahí que existen IJ tratamientos diferentes. El número de observaciones realizadas en el tratamiento (i, j) serán denotadas por K_{ij}. En la sección 11.1 se considera $K_{ij} = 1$. Un caso especial importante de este tipo es un diseño de bloque aleatorizado, en el cual un solo factor A es de primordial interés pero se crea otro factor, "bloques", para controlar la variabilidad externa en unidades o sujetos experimentales. En la sección 11.2 se aborda el caso $K_{ij} = K > 1$ y se mencionan brevemente las dificultades asociadas con K_{ij} desiguales.

La sección 11.3 considera experimentos que implican más de dos factores. Cuando el número de factores es grande, un experimento compuesto de por lo menos una observación por cada tratamiento sería caro y consumiría mucho tiempo. Una situación que se encuentra con frecuencia, la cual se discute en la sección 11.4, es aquella en la que existen p factores, cada uno de los cuales tiene dos niveles. Existen entonces 2^p tratamientos diferentes. Se considera el caso en el cual las observaciones se realizan en todos estos tratamientos (un diseño completo) y el caso en el cual las observaciones se realizan en sólo un subconjunto seleccionado de tratamientos (un diseño incompleto).

11.1 ANOVA bifactorial con $K_{ij} = 1$

Cuando el factor A consta de I niveles y el factor B de J niveles, existen IJ combinaciones diferentes (pares) de niveles de los dos factores, cada uno llamado tratamiento. Con K_{ij} = el número de observaciones en el tratamiento compuesto del factor A al nivel i y del factor B a nivel j, esta sección se enfoca en el caso $K_{ij} = 1$, de modo que los datos se componen de IJ observaciones. Primero se discutirá el modelo de efectos fijos, en el cual los únicos niveles de interés con los dos factores son aquellos que en realidad están representados en el experimento. Situaciones en las que al menos un factor es aleatorio se discuten con brevedad al final de la sección.

Ejemplo 11.1 ¿Es realmente fácil eliminar manchas en telas producidas por plumas de tinta borrable como la palabra *borrable* podría implicar? Considere los siguientes datos de un experimento para comparar tres marcas diferentes de plumas y cuatro tratamientos de lavado diferentes con respecto a su capacidad de eliminar manchas en un tipo particular de tela (basado en "An Assessment of the Effects of Treatment, Time, and Heat on the Removal of Erasable Pen Marks from Cotton and Cotton/Poliester Blend Fabrics", *J. of Testing and Evaluation*, 1991: 394–397). La variable de respuesta es un indicador cuantitativo del cambio de color total de un espécimen de tela; mientras más bajo fue este valor, más manchas fueron eliminadas.

		Tratamiento de lavado					
		1	**2**	**3**	**4**	**Total**	**Promedio**
Marca de pluma	**1**	.97	.48	.48	.46	2.39	.598
	2	.77	.14	.22	.25	1.38	.345
	3	.67	.39	.57	.19	1.82	.455
	Total	2.41	1.01	1.27	.90	5.59	
	Promedio	.803	.337	.423	.300		.466

¿Existe alguna diferencia en la cantidad de cambio de color promedio verdadero debido a las diferentes marcas de pluma o a los diferentes tratamientos de lavado? ■

Como en el ANOVA unifactorial, se utilizan subíndices dobles para identificar variables aleatorias y valores observados. Sea

X_{ij} = la variable aleatoria (va) que denota la medición cuando el factor A se mantiene al nivel i y el factor B al nivel j.

x_{ij} = el valor observado de X_{ij}.

Las x_{ij} normalmente se presentan en una tabla rectangular en la que varios renglones son identificados con los niveles del factor A y varias columnas con los niveles del factor B. En el experimento de la pluma de tinta borrable del ejemplo 11.1, el número de niveles del factor A es $I = 3$, el número de niveles del factor B es $J = 4$, $x_{13} = .48$, $x_{22} = .14$, etcétera.

En tanto en el ANOVA unifactorial, lo único que interesaba eran las medias que aparecían en los renglones y la media grande, en este caso también existe interés en las medias que aparecen en las columnas. Sean

$$\overline{X}_{i\cdot} = \text{el promedio de las mediciones obtenidas cuando el factor } A \text{ se mantiene al nivel } i = \frac{\sum_{j=1}^{J} X_{ij}}{J}$$

$$\overline{X}_{\cdot j} = \text{el promedio de las mediciones obtenidas cuando el factor } B \text{ se mantiene al nivel } j = \frac{\sum_{i=1}^{I} X_{ij}}{I}$$

$$\overline{X}_{\cdot\cdot} = \text{la media grande} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J} X_{ij}}{IJ}$$

con los valores observados $\bar{x}_{i\cdot}$, $\bar{x}_{\cdot j}$, y $\bar{x}_{\cdot\cdot}$. Los totales en lugar de los promedios se denotan omitiendo la raya horizontal (por lo tanto $x_{\cdot j} = \Sigma_i x_{ij}$, etc.). Intuitivamente, para ver si existe algún efecto debido a los niveles del factor A, se deberá comparar las $\bar{x}_{i\cdot}$ observadas una con otra y se deberá sacar información de las $\bar{x}_{\cdot j}$ con respecto a los diferentes niveles del factor B.

El modelo de efectos fijos

Procediendo por analogía con el ANOVA unifactorial, la primera tendencia al especificar un modelo es hacer $\mu_{ij} =$ la respuesta promedio verdadera cuando el factor A se encuentra al nivel i y el factor B al nivel j, para obtener parámetros medios IJ. En ese caso sea

$$X_{ij} = \mu_{ij} + \epsilon_{ij}$$

donde ϵ_{ij} es la cantidad aleatoria en la cual el valor observado difiere de su expectativa y las ϵ_{ij} se suponen normales e independientes con varianza común σ^2. Desafortunadamente, no existe un procedimiento de prueba válido para esta selección de parámetros. Esto es porque hay $IJ + 1$ parámetros (las μ_{ij} y σ^2) pero sólo observaciones IJ, así que después de usar cada x_{ij} como una estimación de μ_{ij}, no hay manera de estimar σ^2.

El modelo siguiente alternativo es realista pero implica relativamente pocos parámetros.

Supóngase la existencia de I parámetros $\alpha_1, \alpha_2, \ldots, \alpha_I$ y J parámetros $\beta_1, \beta_2, \ldots, \beta_J$, de tal suerte que

$$X_{ij} = \alpha_i + \beta_j + \epsilon_{ij} \quad (i = 1, \ldots, I, \quad j = 1, \ldots, J) \tag{11.1}$$

de modo que

$$\mu_{ij} = \alpha_i + \beta_j \tag{11.2}$$

Incluida σ^2, ahora hay $I + J + 1$ parámetros de modelo, así que si $I \geq 3$ y $J \geq 3$, entonces habrá menos parámetros que observaciones (de hecho, en breve se modificará (11.2) de modo que incluso $I = 2$ y/o $J = 2$ tendrán cabida).

El modelo especificado en (11.1) y (11.2) se llama **modelo aditivo** porque cada respuesta media μ_{ij} es la suma de un efecto debido al factor A al nivel i (α_i) y un efecto debido

al factor B al nivel j (β_j). La diferencia entre las respuestas medias con el factor A al nivel i y al nivel i' cuando B se mantiene al nivel j es $\mu_{ij} - \mu_{i'j}$. Cuando el modelo es aditivo,

$$\mu_{ij} - \mu_{i'j} = (\alpha_i + \beta_j) - (\alpha_{i'} + \beta_j) = \alpha_i - \alpha_{i'}$$

la cual es independiente del nivel j del segundo factor. Un resultado similar prevalece para $\mu_{ij} - \mu_{ij'}$. Así pues aditividad significa que la diferencia en las respuestas medias a dos niveles de uno de los factores es la misma a todos los niveles del otro factor. La figura 11.1(a) muestra un conjunto de respuestas medias que satisfacen la condición de aditividad y la figura 11.1(b) muestra una configuración no aditiva.

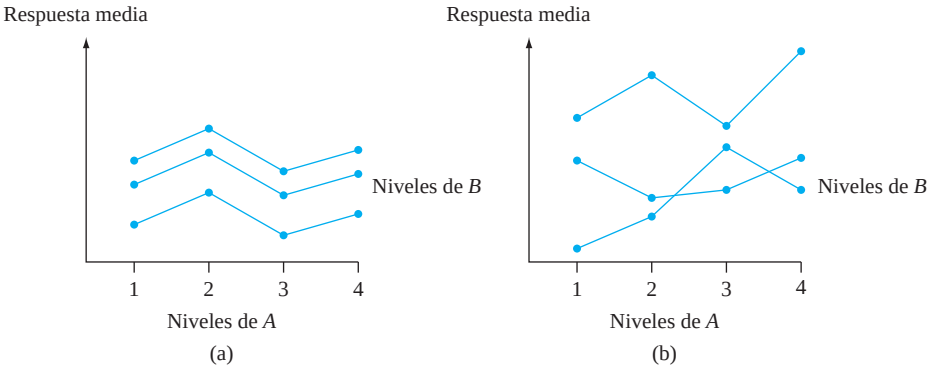


Figura 11.1 Respuestas medias de dos tipos de modelo: (a) aditivo; (b) no aditivo

Ejemplo 11.2 (Continuación del ejemplo 11.1)

Grafique de una manera análoga a la de la figura 11.1 las x_{ij} observadas, para obtener el resultado mostrado en la figura 11.2. Aunque existe algo de "cruzamiento" en las x_{ij} observadas, la configuración es razonablemente representativa de lo que se esperaría bajo aditividad con sólo una observación por tratamiento.

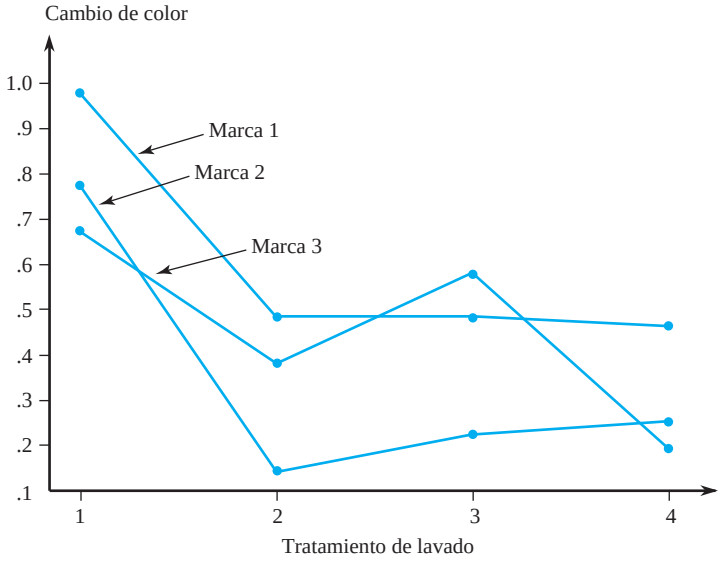


Figura 11.2 Gráfica de los datos del ejemplo 11.1

La expresión (11.2) no describe del todo el modelo final porque las α_i y las β_j no están determinadas de forma única. A continuación se dan configuraciones diferentes de las α_i y β_j que dan las mismas μ_{ij} aditivas.

	$\beta_1 = 1$	$\beta_2 = 4$
$\alpha_1 = 1$	$\mu_{11} = 2$	$\mu_{12} = 5$
$\alpha_2 = 2$	$\mu_{21} = 3$	$\mu_{22} = 6$

	$\beta_1 = 2$	$\beta_2 = 5$
$\alpha_1 = 0$	$\mu_{11} = 2$	$\mu_{12} = 5$
$\alpha_2 = 1$	$\mu_{21} = 3$	$\mu_{22} = 6$

Si se resta cualquier constante c de todas las α_i y se suma a c todas las β_j se obtienen otras configuraciones correspondientes al mismo modelo aditivo. Esta no singularidad se elimina con el uso del siguiente modelo.

$$X_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \tag{11.3}$$

donde $\sum_{i=1}^{I} \alpha_i = 0$, $\sum_{j=1}^{J} \beta_j = 0$, y las ϵ_{ij} se suponen independientes, normalmente distribuidas con media 0 y varianza común σ^2.

Esto es análogo a la selección alternativa de parámetros para el método ANOVA unifactorial discutido en la sección 10.3. No es difícil verificar que (11.3) es un modelo aditivo en el cual los parámetros están determinados de forma única (por ejemplo, para las μ_{ij} previamente mencionadas, $\mu = 4$, $\alpha_1 = -.5$, $\alpha_2 = .5$, $\beta_1 = -1.5$ y $\beta_2 = 1.5$). Obsérvese que hay sólo $I - 1$ α_i independientemente determinadas y $J - 1$ β_j independientemente determinadas, así que (incluida μ) (11.3) especifica $I + J - 1$ parámetros medios.

La interpretación de los parámetros en (11.3) es directa: μ es la media grande verdadera (respuesta media promediada a todos los niveles de ambos factores) α_i es el efecto del factor A al nivel i (medido como una desviación con respecto a μ) y β_j es el efecto del factor B al nivel j. Estimadores insesgados (y la máxima probabilidad) de estos parámetros son

$$\hat{\mu} = \bar{X}_{..} \quad \hat{\alpha}_i = \bar{X}_{i.} - \bar{X}_{..} \quad \hat{\beta}_j = \bar{X}_{.j} - \bar{X}_{..}$$

Existen dos hipótesis nulas diferentes de interés en un experimento de dos factores con $K_{ij} = 1$. La primera, denotada por H_{0A}, establece que los diferentes niveles del factor A no tienen efecto en la respuesta promedio verdadera. La segunda, denotada por H_{0B} asevera que el factor B no tiene ningún efecto.

$$\begin{aligned} &H_{0A}\text{: } \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0 \\ &\qquad \text{contra } H_{aA}\text{: por lo menos una } \alpha_i \neq 0 \\ &H_{0B}\text{: } \beta_1 = \beta_2 = \cdots = \beta_J = 0 \\ &\qquad \text{contra } H_{aB}\text{: por lo menos una } \beta_j \neq 0 \end{aligned} \tag{11.4}$$

(Ningún efecto del factor A implica que todas las α_i son iguales, así que todas deben ser 0 puesto que suman 0 y asimismo para las β_j.)

Procedimientos de prueba

La descripción y análisis ahora siguen de cerca a los del ANOVA unifactorial. Ahora hay cuatro sumas de cuadrados, cada una con un número de grados de libertad asociados:

DEFINICIÓN

$$\text{SST} = \sum_{i=1}^{I}\sum_{j=1}^{J}(X_{ij} - \bar{X}_{..})^2 \qquad \text{gl} = IJ - 1$$

$$\text{SSA} = \sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{X}_{i.} - \bar{X}_{..})^2 = J\sum_{i=1}^{I}(\bar{X}_{i.} - \bar{X}_{..})^2 \quad \text{gl} = I - 1$$

$$\text{SSB} = \sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{X}_{.j} - \bar{X}_{..})^2 = I\sum_{j=1}^{J}(\bar{X}_{.j} - \bar{X}_{..})^2 \quad \text{gl} = J - 1 \qquad (11.5)$$

$$\text{SSE} = \sum_{i=1}^{I}\sum_{j=1}^{J}(X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..})^2 \qquad \text{gl} = (I - 1)(J - 1)$$

La identidad fundamental es

$$\text{SST} = \text{SSA} + \text{SSB} + \text{SSE} \qquad (11.6)$$

Existen fórmulas para SST, SSA y SSB análogas a las que aparecen en el capítulo 10 para el ANOVA unifactorial. Pero la amplia disponibilidad de programas estadísticos ha vuelto a estas fórmulas casi obsoletas.

La expresión para SSE se obtiene al reemplazar μ, α_i y β_j por sus estimadores en $\sum[X_{ij} - (\mu + \alpha_i + \beta_j)]^2$. El grado de libertad asociado con el error es IJ − número de parámetros medios estimado $= IJ - [1 + (I - 1) + (J - 1)] = (I - 1)(J - 1)$. La variación total se divide en una parte (SSE) que no está explicada por la verdad o falsedad de H_{0A} o H_{0B} y dos partes que pueden ser explicadas por la posible falsedad de las dos hipótesis nulas.

La teoría estadística ahora estipula que si se forman proporciones F como en el ANOVA unifactorial, cuando H_{0A} (H_{0B}) es verdadera, la proporción F correspondiente tiene una distribución F con grados de libertad asociados con el numerador $= I - 1$ $(J - 1)$ y grados de libertad asociados con el denominador $= (I - 1)(J - 1)$.

Hipótesis	Valor del estadístico de prueba	Región de rechazo
H_{0A} contra H_{aA}	$f_A = \dfrac{\text{MSA}}{\text{MSE}}$	$f_A \geq F_{\alpha,I-1,(I-1)(J-1)}$
H_{0B} contra H_{aB}	$f_B = \dfrac{\text{MSB}}{\text{MSE}}$	$f_B \geq F_{\alpha,J-1,(I-1)(J-1)}$

Ejemplo 11.3 (Continuación del ejemplo 11.2)

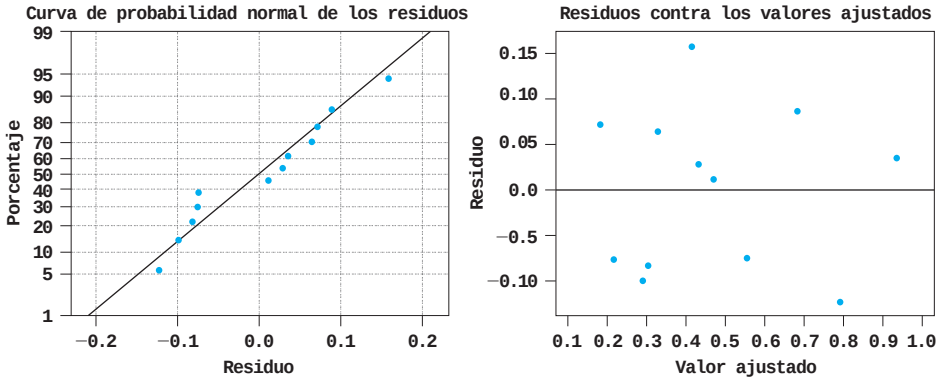
Las $\bar{x}_{i.}$ y $\bar{x}_{.j}$ para los datos de cambio de color se muestran a lo largo de los márgenes de la tabla de datos dada previamente. La tabla 11.1 resume los cálculos.

Tabla 11.1 Tabla ANOVA para el ejemplo 11.3

Causa de la variación	Grados de libertad	Suma de cuadrados	Media cuadrática	f
Factor A (marca)	$I - 1 = 2$	SSA = .1282	MSA = .0641	$f_A = 4.43$
Factor B (tratamiento de lavado)	$J - 1 = 3$	SSB = .4797	MSB = .1599	$f_B = 11.05$
Error	$(I - 1)(J - 1) = 6$	SSE = .0868	MSE = .01447	
Total	$IJ - 1 = 11$	SST = .6947		

El valor crítico para probar H_{0A} a un nivel de significancia de .05 es $F_{.05,2,6} = 5.14$. Como $4.43 < 5.14$, H_{0A} no puede ser rechazada a un nivel de significancia de .05. Aparentemente el cambio de color promedio verdadero no depende de la marca de la pluma. Como $F_{.05,3,6} = 4.76$ y $11.05 \geq 4.76$, H_{0B} es rechazada a un nivel de significancia de .05 a favor de la aseveración de que el cambio de color varía con el tratamiento de lavado. Un programa estadístico da valores P de .066 y .007 con estas dos pruebas. ■

La plausibilidad de las suposiciones de normalidad y varianza constante puede ser investigada gráficamente. Se definen los *valores pronosticados* (también llamados *valores ajustados*) $\hat{x}_{ij} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j = \bar{x}_{..} + (\bar{x}_{i.} - \bar{x}_{..}) + (\bar{x}_{.j} - \bar{x}_{..}) = \bar{x}_{i.} + \bar{x}_{.j} - \bar{x}_{..}$ y los residuos (las diferencias entre las observaciones y los valores pronosticados) $x_{ij} - \hat{x}_{ij} = x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..}$. Se puede verificar la suposición de normalidad con una gráfica de probabilidad normal de los residuos y la suposición de varianza constante con una gráfica de los residuos contra los valores ajustados. La figura 11.3 muestra estas gráficas para los datos del ejemplo 11.3.



(a) (b)

Figura 11.3 Gráficas de diagnóstico obtenidas con MINITAB para el ejemplo 11.3

La gráfica de probabilidad normal es razonablemente lineal, de modo que no existe ninguna razón para cuestionar la normalidad de este conjunto de datos. En la gráfica de residuos contra valores ajustados, se buscan diferencias en dispersión vertical conforme se recorre la gráfica horizontalmente de un lado a otro. Por ejemplo, si hubiera un rango angosto de valores ajustados pequeños y un rango amplio de valores ajustados altos, esto sugeriría que la varianza es más grande con respuestas grandes (esto sucede a menudo y en ocasiones puede ser remediado reemplazando cada observación por su logaritmo). La gráfica 11.3 (b) muestra que no existe evidencia en contra de la suposición de varianza constante.

Media cuadrática esperada

La plausibilidad de utilizar las pruebas F que se acaban de describir se demuestra calculando la media cuadrática esperada. Para el modelo aditivo,

$$E(\text{MSE}) = \sigma^2$$

$$E(\text{MSA}) = \sigma^2 + \frac{J}{I-1}\sum_{i=1}^{I}\alpha_i^2$$

$$E(\text{MSB}) = \sigma^2 + \frac{I}{J-1}\sum_{j=1}^{J}\beta_j^2$$

Si H_{0A} es verdadera, MSA es un estimador insesgado de σ^2, así que F es una razón de dos estimadores de σ^2 insesgados. Cuando H_{0A} es falsa, MSA tiende a sobrestimar σ^2, de tal suerte que H_{0A} deberá ser rechazada cuando la proporción F_A es demasiado grande. Comentarios similares se aplican a MSB y H_{0B}.

Comparaciones múltiples

Cuando H_{0A} o H_{0B} ha sido rechazada, se puede utilizar el procedimiento de Tukey para identificar diferencias significativas entre los niveles del factor investigado.

1. Para comparar niveles del factor A, se obtiene $Q_{\alpha,I,(I-1)(J-1)}$.

Para comparar niveles del factor B, se obtiene $Q_{\alpha,J,(I-1)(J-1)}$.

2. Se calcula

$$w = Q \cdot (\text{desviación estándar estimada de las medias muestrales comparadas})$$

$$= \begin{cases} Q_{\alpha,I,(I-1)(J-1)} \cdot \sqrt{\text{MSE}/J} & \text{para comparaciones del factor } A \\ Q_{\alpha,J,(I-1)(J-1)} \cdot \sqrt{\text{MSE}/I} & \text{para comparaciones del factor } B \end{cases}$$

(porque, p. ej., la desviación estándar de $\bar{X}_{i\cdot}$ es $\sigma/\sqrt{J}$).

3. Se ordenan las medias muestrales en orden creciente, se subrayan los pares que difieren por menos de w y se identifican los pares no subrayados por la misma línea como correspondientes a niveles significativamente diferentes del factor dado.

Ejemplo 11.4 (Continuación del ejemplo 11.3)

La identificación de diferencias significativas entre los cuatro tratamientos de lavado requiere $Q_{.05,4,6} = 4.90$ y $w = 4.90\sqrt{(.01447)/3} = .340$. Las cuatro medias muestrales correspondientes al factor B (promedios en columnas) ahora aparecen listadas en orden creciente y cualquier par que difiere por menos de .340 aparece subrayado por un segmento de recta:

$\bar{x}_{4\cdot}$	$\bar{x}_{2\cdot}$	$\bar{x}_{3\cdot}$	$\bar{x}_{1\cdot}$
.300	337	.423	.803

Parece que el tratamiento de lavado 1 difiere significativamente de los otros tres, aunque no están identificadas ningunas otras diferencias importantes. En particular, no es aparente cuál entre los tratamientos 2, 3 y 4 es mejor para eliminar manchas. ■

Experimentos de bloque aleatorizado

Cuando se utiliza el ANOVA unifactorial para probar en cuanto a la presencia de efectos debidos a los I tratamientos diferentes estudiados, una vez que los IJ sujetos o unidades experimentales han sido seleccionados, el tratamiento se asignará en una forma completamente al azar. Es decir, se deberán seleccionar al azar J sujetos para el primer tratamiento, luego otra muestra de J sujetos seleccionados al azar de los $IJ - J$ sujetos restantes para el segundo tratamiento y así sucesivamente.

Sucede con frecuencia, no obstante, que los sujetos o unidades experimentales exhiben heterogeneidad con respecto a otras variables que pueden afectar las respuestas observadas. Cuando éste es el caso, la presencia o ausencia de un valor F significativo puede deberse a esta variación externa y no a la presencia o ausencia de efectos factoriales. Ésta fue la razón para la introducción de experimentos apareados en el capítulo 9. La analogía con un experimento apareado cuando $I > 2$ se llama experimento de **bloque aleatorizado.** Un factor externo "bloques", se construye dividiendo las IJ unidades en J grupos con I uni-

dades en cada grupo. Este agrupamiento o formación de bloques se realiza de tal modo que dentro de cada bloque las I unidades son homogéneas con respecto a otros factores que se piensa afectan las respuestas. Entonces dentro de cada bloque homogéneo, los I tratamientos se asignan al azar a las I unidades o sujetos.

Ejemplo 11.5 Una organización de prueba de productos de consumo deseaba comparar el consumo de energía anual de cinco marcas diferentes de deshumidificadores. Como el consumo de energía depende del nivel de humedad prevaleciente, se decidió monitorear cada marca a cuatro niveles diferentes desde humedad moderada hasta intensa (formando así un bloque con el nivel de humedad). Dentro de cada nivel, se asignaron las marcas al azar a los cinco lugares seleccionados. Las observaciones resultantes (kWh anuales) aparecen en la tabla 11.2 y los cálculos ANOVA se resumen en la tabla 11.3.

Tabla 11.2 Datos de consumo de energía del ejemplo 11.5

Tratamientos (marcas)	Bloques (nivel de humedad) 1	2	3	4	$x_{i\cdot}$	$\bar{x}_{i\cdot}$
1	685	792	838	875	3190	797.50
2	722	806	893	953	3374	843.50
3	733	802	880	941	3356	839.00
4	811	888	952	1005	3656	914.00
5	828	920	978	1023	3749	937.25
$x_{\cdot j}$	3779	4208	4541	4797	17,325	
$\bar{x}_{\cdot j}$	755.80	841.60	908.20	959.40		866.25

Tabla 11.3 Tabla ANOVA para el ejemplo 11.5

Causa de la variación	Grados de libertad	Suma de cuadrados	Media cuadrática	f
Tratamientos (marcas)	4	53,231.00	13,307.75	$f_A = 95.57$
Bloques	3	116,217.75	38,739.25	$f_B = 278.20$
Error	12	1671.00	139.25	
Total	19	171,119.75		

Como $F_{.05,4,12} = 3.26$ y $f_A = 95.57 \geq 3.26$, H_0 es rechazada a favor de H_a y se concluye que el consumo de energía si depende de la marca del humidificador. Para identificar marcas significativamente diferentes, se utiliza el procedimiento de Tukey, $Q_{.05,5,12} = 4.51$ y $w = 4.51\sqrt{139.25/4} = 26.6$.

$\bar{x}_{1\cdot}$	$\bar{x}_{3\cdot}$	$\bar{x}_{2\cdot}$	$\bar{x}_{4\cdot}$	$\bar{x}_{5\cdot}$
797.50	839.00	843.50	914.00	937.25

El subrayado indica que las marcas pueden dividirse en tres grupos con respecto a consumo de energía.

Como el factor de bloque es de interés secundario, no se requiere $F_{.05,3,12}$, aun cuando el valor calculado de F_B es claramente muy significativo. La figura 11.4 muestra los resultados generados por SAS con estos datos. Obsérvese que en la primera parte de la tabla ANOVA, las sumas de los cuadrados (SS) para tratamientos (marcas) y bloques (niveles de humedad) se combinan en una sola suma de los cuadrados "modelo" SS.

```
                              Análisis de procedimiento de varianza
Variable dependiente: USO DE ENERGÍA
                                      Suma de            Media
Fuente                      DF        cuadrados          cuadrática          Valor F          Pr > F
Modelo                       7       169448.750         24206.964            173.84          0.0001
Error                       12         1671.000           139.250
Corregido Total             19       171119.750

                                                                          Media de
                  R-Cuadrado          C.V.              Raíz MSE        USO DE ENERGÍA
                    0.990235         1.362242            11.8004          866.25000

Fuente                 DF          Anova SS       Media cuadrática        Valor F          PR > F
MARCA                   4         53231.000          13307.750             95.57          0.0001
HUMEDAD                 3        116217.750          38739.250            278.20          0.0001

          Alfa = 0.05  df = 12  MSE = 139.25
       Valor crítico de rango estudentizado = 4.508
         Diferencia significativa mínima = 26.597

Las medias con la misma letra no son significativamente diferentes.

       Grupo Tukey              Media          N          MARCA
                A             937.250          4              5
                A
                A             914.000          4              4
                B             843.500          4              2
                B
                B             839.000          4              3
                C             797.500          4              1
```

Figura 11.4 Resultados obtenidos con SAS con los datos de consumo de energía ■

En muchas situaciones experimentales en las que los tratamientos tienen que ser aplicados a sujetos, uno solo de ellos puede recibir todos los I tratamientos. La formación de bloques se realiza entonces con los sujetos mismos para controlar la variabilidad entre ellos; se dice entonces que cada sujeto actúa como su propio control. Los científicos sociales en ocasiones se refieren a tales experimentos como diseños de medidas repetidas. Las "unidades" dentro de un bloque son entonces las diferentes "instancias" de aplicación de tratamiento. Del mismo modo, los bloques se consideran como lapsos de tiempo, ubicaciones u observadores diferentes.

Ejemplo 11.6 ¿Cómo afecta la tensión de las cuerdas de las raquetas de tenis la velocidad de la pelota que sale de la raqueta? El artículo "Elite Tennis Player Sensitivity to Changes in String Tension and the Effect on Resulting Ball Dynamics" (*Sports Engr.*, 2008: 31–36) describe un experimento en el que cuatro diferentes tensiones de cuerda (N) fueron utilizadas y las bolas proyectadas desde una máquina fueron golpeadas por 18 jugadores diferentes. La velocidad de rebote (km/h) se determinó para cada combinación tensión-jugador. Considere los siguientes datos en la tabla 11.4 de un experimento similar con sólo seis jugadores (el resultado de ANOVA se encuentra en buen acuerdo con lo reportado en el artículo).

Los cálculos ANOVA de análisis de varianza se resumen en la tabla 11.5. El valor P para la prueba para ver si la velocidad promedio real de rebote depende de la tensión de las cuerdas es .049. Por lo tanto H_0: $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ es apenas rechazada al nivel de significancia .05 a favor de la conclusión de que la velocidad promedio real varía con la tensión ($F_{.05,3,15} = 3.29$). La aplicación del procedimiento de Tukey para identificar diferencias significativas entre las tensiones requiere $Q_{.05,4,15} = 4.08$. Entonces $w = 7.464$. La diferencia entre la media de la muestra mayor y menor de las tensiones es 6.87. Así, aunque la prueba F es significativa, el método de Tukey no identifica ninguna. En ocasiones,

Tabla 11.4 Datos de velocidad de rebote del ejemplo 11.6

	Jugador						
Tensión	**1**	**2**	**3**	**4**	**5**	**6**	$\bar{x}_{i\cdot}$
210	105.7	116.6	106.6	113.9	119.4	123.5	114.28
235	113.3	119.9	120.5	119.3	122.5	124.0	119.92
260	117.2	124.4	122.3	120.0	115.1	127.9	121.15
285	110.0	106.8	110.0	115.3	122.6	128.3	115.50
$\bar{x}_{\cdot j}$	111.55	116.93	114.85	117.13	119.90	125.93	

Tabla 11.5 Tabla ANOVA del ejemplo 11.6

Fuente	**gl**	**SS**	**MS**	***f***	***P***
Tensión	3	199.975	66.6582	3.32	0.049
Jugador	5	477.464	95.4928	4.76	0.008
Error	15	301.188	20.0792		
Total	23	978.626			

esto sucede cuando la hipótesis nula es apenas rechazada. La configuración de la media muestral en el artículo citado es similar a la nuestra. Los autores comentan que los resultados eran contrarios a las anteriores pruebas de laboratorio, donde el aumento de las velocidades de rebote están típicamente asociadas con baja tensión de las cuerdas. ■

En la mayoría de los experimentos de bloques aleatorizados en los cuales los sujetos se desempeñan como bloques, los sujetos que en realidad participan en el experimento se seleccionan de una gran población. Los sujetos contribuyen entonces con efectos aleatorios en lugar de fijos. Esto no afecta el procedimiento de comparar tratamientos cuando $K_{ij} = 1$ (una observación por "celda" como en esta sección), pero el procedimiento cambia si $K_{ij} = K > 1$. En breve se considerarán modelos de dos factores en los cuales los efectos son aleatorios.

Más sobre formación de bloques Cuando $I = 2$, se puede utilizar la prueba F o la prueba t de diferencias apareadas para analizar los datos. La conclusión resultante no dependerá de cuál procedimiento se utilice, puesto que $T^2 = F$ y $t^2_{\alpha/2,\nu} = F_{\alpha,1,\nu}$.

Al igual que con la formación de pares, la formación de bloques implica tanto una ganancia como una pérdida potencial de precisión. Si existe una gran cantidad de heterogeneidad en las unidades experimentales, el valor del parámetro de varianza σ^2 en el modelo unidireccional será grande. El efecto de la formación de bloques es filtrar la variación representada por σ^2 en el modelo bidireccional apropiado para un experimento de bloques aleatorizados. Con las demás cosas iguales, un valor de σ^2 más pequeño da por resultado una prueba que es más probable que detecte alejamientos de H_0 (es decir, una prueba con mayor potencia).

Sin embargo, las demás cosas no son iguales aquí, puesto que la prueba F unifactorial está basada en $I(J - 1)$ grados de libertad (gl) para el error, mientras que la prueba F bifactorial está basada en $(I - 1)(J - 1)$ grados de libertad en el caso de error. Pocos grados de libertad en el caso de error reducen la potencia, en esencia porque el estimador asociado con el denominador de σ^2 no es tan preciso. Esta pérdida de grados de libertad puede ser especialmente seria si el experimentador sólo puede permitirse un pequeño número de observaciones. No obstante, si aparece la formación de bloques reduce significativamente la variabilidad, el sacrificio del grado de libertad del error es sensible.

Modelos de efectos aleatorios y combinados

En muchos experimentos, los niveles reales de un factor utilizados en el experimento, y no los que interesan al experimentador, se seleccionan de una población mucho más grande de niveles posibles del factor. Si esto es cierto para ambos factores en un experimento bifactorial, un **modelo de efectos aleatorios** es apropiado. El caso en el cual los niveles de un factor son los únicos de interés y los niveles del otro factor se seleccionan de una población de niveles conduce a un **modelo de efectos combinados.** El modelo de efectos aleatorios bifactorial cuando $K_{ij} = 1$ es

$$X_{ij} = \mu + A_i + B_j + \epsilon_{ij} \quad (i = 1, \ldots, I, \quad j = 1, \ldots, J)$$

Las A_i, B_j y ϵ_{ij} son variables aleatorias independientes normalmente distribuidas con media 0 y varianzas σ_A^2, σ_B^2 y σ^2, respectivamente. Las hipótesis de interés son entonces H_{0A}: $\sigma_A^2 = 0$ (el nivel del factor A no contribuye a la variación de la respuesta) contra H_{aA}: $\sigma_A^2 > 0$ y H_{0B}: $\sigma_B^2 = 0$ contra H_{aB}: $\sigma_B^2 > 0$. En tanto que $E(\text{MSE}) = \sigma^2$ como antes, los cuadrados medios esperados para los factores A y B ahora son

$$E(\text{MSA}) = \sigma^2 + J\sigma_A^2 \qquad E(\text{MSB}) = \sigma^2 + I\sigma_B^2$$

Por consiguiente cuando H_{0A} (H_{0B}) es verdadera, F_A(F_B) sigue siendo la razón de dos estimadores insesgados de σ^2. Se puede demostrar que una prueba a nivel α para H_{0A} contra H_{aA} rechaza H_{0A} si $f_A \geq F_{\alpha,I-1,(I-1)(J-1)}$ y, asimismo, se utiliza el mismo procedimiento que antes para decidir entre H_{0B} y H_{aB}.

Para el caso en el cual el factor A es fijo y el B es aleatorio, el modelo combinado es

$$X_{ij} = \mu + \alpha_i + B_j + \epsilon_{ij} \quad (i = 1, \ldots, I, \quad j = 1, \ldots, J)$$

donde $\sum \alpha_i = 0$ y las B_j y ϵ_{ij} están normalmente distribuidas con media 0 y varianzas σ_B^2 y σ^2, respectivamente. Ahora las dos hipótesis nulas son

$$H_{0A}: \alpha_1 = \cdots = \alpha_I = 0 \text{ y } H_{0B}: \sigma_B^2 = 0$$

con medias cuadráticas esperadas

$$E(\text{MSE}) = \sigma^2 \qquad E(\text{MSA}) = \sigma^2 + \frac{J}{I-1}\sum \alpha_i^2 \qquad E(\text{MSB}) = \sigma^2 + I\sigma_B^2$$

Los procedimientos de prueba para H_{0A} contra H_{aA} y H_{0B} contra H_{aB} son exactamente como antes. Por ejemplo, en el análisis de los datos de cambio de color en el ejemplo 11.1, si los cuatro tratamientos de lavado fueron seleccionados al azar, entonces como $f_B = 11.05$ y $F_{.05,3,6} = 4.76$, H_{0B}: $\sigma_B^2 = 0$ es rechazada a favor de H_{aB}: $\sigma_B^2 > 0$. Entonces (MSB − MSE)/I = .0485 da una estimación del "componente de varianza" σ_B^2.

Resumiendo, cuando $K_{ij} = 1$, aunque las hipótesis y los cuadrados medios esperados difieren con ambos efectos fijos, los procedimientos de prueba son idénticos.

EJERCICIOS Sección 11.1 (1–15)

1. Se determinó el número de millas (en miles) de desgaste útil de la banda de rodamiento de llantas de cada una de cinco marcas diferentes de carros subcompactos (factor A, con $I = 5$) en combinación con cada una de cuatro marcas diferentes de llantas radiales (factor B, con $J = 4$) y se obtuvieron $IJ = 20$ observaciones. Se calcularon entonces los valores SSA = 30.6, SSB = 44.1 y SSE = 59.2. Suponga que un modelo aditivo es apropiado.

a. Pruebe H_0: $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ (no hay diferencias en la vida útil de las llantas promedio verdadera a causa de las marcas de los carros) contra H_a: por lo menos $\alpha_i \neq 0$ con una prueba al nivel de .05.

b. H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ (no hay diferencias en la vida útil de las llantas promedio verdadera debido a las marcas de las llantas) contra H_a: por lo menos una $\beta_j \neq 0$ utilizando una prueba al nivel de .05.

2. Se están considerando cuatro recubrimientos diferentes como protección contra corrosión de tubería de metal. Ésta se enterrará

en tres tipos diferentes de suelo. Para investigar si la cantidad de corrosión depende del recubrimiento o del tipo de suelo, se seleccionan 12 tramos de tubería. Cada tramo se recubre con uno de los cuatro recubrimientos y se entierra en uno de los tres tipos de suelo durante un tiempo fijo, después del cual se determina la cantidad de corrosión (profundidad máxima de las picaduras, en .0001 pulg). Los datos aparecen en la tabla.

		Tipo de suelo (*B*)		
		1	**2**	**3**
Recubrimiento (*A*)	**1**	64	49	50
	2	53	51	48
	3	47	45	50
	4	51	43	52

a. Suponiendo la validez del modelo aditivo realice el análisis ANOVA por medio de una tabla ANOVA para ver si la cantidad de corrosión depende del tipo de recubrimiento utilizado o del tipo de suelo. Use $\alpha = .05$.

b. Calcule $\hat{\mu}$, $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$, $\hat{\alpha}_4$, $\hat{\beta}_1$, $\hat{\beta}_2$ y $\hat{\beta}_3$.

3. El artículo "Adiabatic Humidification of Air with Water in a Packed Tower" (*Chem. Eng. Prog.*, 1952: 362–370) reporta datos sobre el coeficiente de transferencia de calor de una película de gas (Btu/h pie^2 en °F) como una función del gasto de gas (factor *A*) y gasto de líquido (factor *B*).

		B			
		1(190)	**2(250)**	**3(300)**	**4(400)**
A	**1(200)**	200	226	240	261
	2(400)	278	312	330	381
	3(700)	369	416	462	517
	4(1100)	500	575	645	733

a. Después de construir una tabla ANOVA, pruebe al nivel .01 con la hipótesis de ningún efecto del gasto de gas contra la alternativa apropiada y la hipótesis de ningún efecto del gasto de líquido contra la alternativa apropiada.

b. Use el procedimiento de Tukey para investigar diferencias en el coeficiente de transferencia de calor esperado debido a los diferentes gastos de gas.

c. Repita el inciso (b) con gastos de líquido.

4. En un experimento para ver si la cantidad de cobertura de pintura de látex de color azul claro para interiores depende de la marca de la pintura o de la marca del rodillo utilizado, se aplicó 1 galón de cada una de las cuatro marcas de pintura utilizando cada una de tres marcas de rodillo y se obtuvieron los siguientes datos (número de pies cuadrados cubiertos).

		Marca de rodillo		
		1	**2**	**3**
Marca de pintura	**1**	454	446	451
	2	446	444	447
	3	439	442	444
	4	444	437	443

a. Construya la tabla ANOVA. [*Sugerencia*: Los cálculos se pueden acelerar restando 400 (o cualquier otro número conveniente) de cada observación. Esto no afecta los resultados finales.]

b. Formule y pruebe las hipótesis apropiadas para decidir si la marca de la pintura tiene algún efecto en la cobertura. Use $\alpha = .05$.

c. Repita el inciso (b) para la marca de rodillo.

d. Use el método de Tukey para identificar diferencias significativas entre las marcas. ¿Hay alguna marca que parezca claramente preferible a las demás?

5. En un experimento para evaluar el efecto de ángulo de tirón de la fuerza requerida para separar conectores eléctricos, se utilizaron cuatro ángulos diferentes (factor *A*) y cada uno de una muestra de cinco conectores (factor *B*) fue jalado una vez a cada ángulo ("A Mixed Model Factorial Experiment in Testing Electrical Connectors", *Industrial Quality Control*, 1960: 12–16). Los datos aparecen en la tabla adjunta.

		B				
		1	**2**	**3**	**4**	**5**
A	**0°**	45.3	42.2	39.6	36.8	45.8
	2°	44.1	44.1	38.4	38.0	47.2
	4°	42.7	42.7	42.6	42.2	48.9
	6°	43.5	45.8	47.9	37.9	56.4

¿Sugieren los datos que la fuerza de separación promedio verdadera es afectada por el ángulo de tirón? Formule y pruebe las hipótesis apropiadas a un nivel de .01 construyendo primero una tabla ANOVA (SST = 396.13, SSA = 58.16 y SSB = 246.97).

6. Un condado particular emplea tres valuadores que son responsables de determinar el valor de las propiedades residenciales en el condado. Para ver si estos valuadores difieren sistemáticamente en sus avalúos, se seleccionan 5 casas y a cada valuador se le pide que determine el valor de mercado de cada casa. Con el factor *A* denotando valuadores ($I = 3$) y el factor *B* denotando casas ($J = 5$), suponga SSA = 11.7, SSB = 113.5 y SSE = 25.6.

a. Pruebe H_0: $\alpha_1 = \alpha_2 = \alpha_3 = 0$ al nivel .05 (H_0 manifiesta que no existen diferencias sistemáticas entre los valuadores.)

b. Explique por qué se utilizó un experimento de bloques aleatorizados con sólo 5 casas en lugar de un experimento ANOVA unidireccional que implique un total de 15 casas diferentes con 5 casas diferentes valuadas por cada asesor (un grupo diferente de 5 para cada valuador).

7. El artículo "Rate of Stuttering Adaptation Under Two Electroshock Conditions" (*Behavior Research Therapy*, 1967, 49–54) da calificaciones de adaptación de tres tratamientos diferentes: (1) ningún choque eléctrico, (2) choque eléctrico después de cada palabra tartamudeada y (3) choque eléctrico durante cada momento de tartamudeo. Estos tratamientos se utilizaron en cada uno de 18 tartamudos y se obtuvo SST = 3476.00, SSTr = 28.78 y SSBI = 2977.67.

a. Construya la tabla ANOVA y pruebe a un nivel de .05 para ver si la calificación de adaptación promedio real depende del tratamiento dado.

b. Juzgando por la proporción F de sujetos (factor B), ¿piensa que la formación de bloques de sujetos fue efectiva en este experimento? Explique.

8. El artículo "Exercise Thermoregulation and Hiperprolactinaemia" (*Ergonomics*, 2005: 1547–1557) discutió cómo varios aspectos de la capacidad de hacer ejercicio podrían depender de la temperatura del ambiente. Los datos adjuntos sobre pérdida de masa corporal (kg) después de ejercitarse en un ergómetro de ciclos semirrecostado en tres diferentes temperaturas ambiente (6°, 18° y 30°C) fueron proporcionados por los autores del artículo.

		Frío	Neutro	Caliente
Sujeto	1	.4	1.2	1.6
	2	.4	1.5	1.9
	3	1.4	.8	1.0
	4	.2	.4	.7
	5	1.1	1.8	2.4
	6	1.2	1.0	1.6
	7	.7	1.0	1.4
	8	.7	1.5	1.3
	9	.8	.8	1.1

a. ¿Afecta la temperatura la pérdida de masa corporal promedio verdadera? Realice una prueba usando un nvel de significancia de .01 (como hicieron los autores en el artículo).

b. Investigue las diferencias significativas entre las temperaturas.

c. Los residuos son .20, .30, −.40, −.07, .30, .00, .03, −.20, −.14, .13, .23, −.27, −.04, .03, −.27, −.04, .33, −.10, −.33 , −.53, .67, .11, −.33, .27, .01, −.13, .24. Úselos como base para investigar la plausibilidad de las suposiciones que fundamentan su análisis en (a).

9. El artículo "The Effects of a Pneumatic Stool and a One-Legged Stool on Lower Limb Joint Load and Muscular Activity During Sitting and Rising" (*Ergonomics*, 1993: 519–535) da los datos adjuntos sobre el esfuerzo requerido de un sujeto para ponerse de pie de cuatro tipos diferentes de bancos (escala de Borg). Analice la varianza con $\alpha = .05$ y continúe con un análisis de comparaciones múltiples si es apropiado.

		Sujeto									
		1	2	3	4	5	6	7	8	9	$\bar{x}_{i\cdot}$
Tipo de banco	1	12	10	7	7	8	9	8	7	9	8.56
	2	15	14	14	11	11	11	12	11	13	12.44
	3	12	13	13	10	8	11	12	8	10	10.78
	4	10	12	9	9	7	10	11	7	8	9.22

10. La resistencia de concreto utilizado en construcciones comerciales tiende a variar de un lote a otro. Por consiguiente, se "curan" pequeños cilindros de prueba de concreto muestreado de un lote durante periodos de hasta 28 días en ambientes con temperatura y humedad controladas antes de realizar mediciones de resistencia. El concreto es entonces "comprado y vendido con base en los cilindros para prueba de resistencia" (ASTM C 31 Standard Test Method for Making and Curing Concrete Test Specimens in the Field). Se obtuvieron los datos adjuntos con un experimento realizado para comparar tres métodos de curado diferentes con respecto a resistencia a la compresión (MPa). Analice estos datos.

Lote	Método A	Método B	Método C
1	30.7	33.7	30.5
2	29.1	30.6	32.6
3	30.0	32.2	30.5
4	31.9	34.6	33.5
5	30.5	33.0	32.4
6	26.9	29.3	27.8
7	28.2	28.4	30.7
8	32.4	32.4	33.6
9	26.6	29.5	29.2
10	28.6	29.4	33.2

11. Para los datos del ejemplo 11.5, compruebe la verosimilitud de las suposiciones mediante la construcción de una gráfica de probabilidad normal de los residuales y una gráfica de los residuos contra los valores pronosticados y comente sobre lo aprendido.

12. Suponga que en el experimento descrito en el ejercicio 6 las cinco casas en realidad se seleccionaron al azar de entre aquellas de una cierta edad y tamaño, de modo que el factor B es aleatorio y no fijo. Pruebe H_0: $\sigma_B^2 = 0$ contra H_a: $\sigma_B^2 > 0$ utilizando una prueba a nivel de .01.

13. **a.** Demuestre que una constante d puede ser sumada a (o restada de) cada x_{ij} sin afectar cualquiera de las sumas de los cuadrados ANOVA.

b. Suponga que cada x_{ij} se multiplica por una constante no cero c. ¿Cómo afecta esto las sumas de los cuadrados ANOVA? ¿Cómo afecta esto los valores de los estadísticos de F, F_A y F_B? ¿Qué efecto tiene la "codificación" de los datos mediante $y_{ij} = cx_{ij} + d$ en las conclusiones que se derivan de los procedimientos ANOVA?

14. Use el hecho de que $E(X_{ij}) = \mu + \alpha_i + \beta_j$ con $\sum\alpha_i = \sum\beta_j = 0$ para demostrar que $E(\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot}) = \alpha_i$, de modo que $\hat{\alpha}_i = \bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot}$ sea un estimador insesgado para α_i.

15. Las curvas de potencia de las figuras 10.5 y 10.6 pueden ser utilizadas para obtener $\beta = P$(error de tipo II) para la prueba F en ANOVA bifactorial. Con valores fijos de $\alpha_1, \alpha_2, \ldots, \alpha_I$, se calcula la cantidad $\phi^2 = (J/I)\sum\alpha_i^2/\sigma^2$. Entonces la cifra correspondiente a $\nu_1 = I - 1$ se ingresa en el eje horizontal en el valor de ϕ, se lee la potencia en el eje vertical de la curva marcada $\nu_2 = (I - 1)(J - 1)$ y $\beta = 1 -$ potencia.

a. En el experimento de corrosión descrito en el ejercicio 2, determine β cuando $\alpha_1 = 4$, $\alpha_2 = 0$, $\alpha_3 = \alpha_4 = -2$ y $\sigma = 4$. Repita con $\alpha_1 = 6$, $\alpha_2 = 0$, $\alpha_3 = \alpha_4 = -3$ y $\sigma = 4$.

b. Por simetría, ¿cuál es β para la prueba H_{0B} contra H_{aB} en el ejemplo 11.1 cuando $\beta_1 = .3$, $\beta_2 = \beta_3 = \beta_4 = -.1$ y $\sigma = .3$?

11.2 ANOVA bifactorial con $K_{ij} > 1$

En la sección 11.1, se analizaron datos obtenidos de un experimento de dos factores en el cual había una observación por cada una de las IJ combinaciones de niveles de los factores. Se supuso que las μ_{ij} tienen una estructura aditiva con $\mu_{ij} = \mu + \alpha_i + \beta_j$, $\sum\alpha_i = \sum\beta_j = 0$. Aditividad significa que la diferencia en las respuestas promedio verdaderas para dos niveles cualesquiera de los factores es la misma para cada nivel del otro factor. Por ejemplo, $\mu_{ij} - \mu_{i'j} = (\mu + \alpha_i + \beta_j) - (\mu + \alpha_{i'} + \beta_j) = \alpha_i - \alpha_{i'}$, independiente del nivel j del segundo factor. Esto se muestra en la figura 11.1(a), donde las líneas que conectan respuestas promedio verdaderas son paralelas.

La figura 11.1(b) ilustra un conjunto de respuestas promedio verdaderas que no tienen estructura aditiva. Las líneas que conectan estas μ_{ij} no son paralelas, lo que significa que la diferencia en las respuestas promedio verdaderas para diferentes niveles de un factor sí dependen del nivel del otro factor. Cuando la aditividad no prevalece, se dice que hay **interacción** entre los diferentes niveles de los factores. La suposición de aditividad permitió en la sección 11.1 obtener un estimador de la varianza del error aleatorio σ^2 (MSE) que resultó ser insesgado fuera o no verdadera cualquier hipótesis nula de interés. Cuando $K_{ij} > 1$ para por lo menos un par (i, j), se puede obtener un estimador válido de σ^2 sin suponer aditividad. Se abordará el caso $K_{ij} = K > 1$, de modo que el número de observaciones por "celda" (por cada combinación de niveles) es constante.

Parámetros de efectos fijos e hipótesis

En lugar de utilizar las μ_{ij} como parámetros de modelo, se acostumbra utilizar un conjunto equivalente que revela con más claridad el rol de interacción.

NOTACIÓN

$$\mu = \frac{1}{IJ}\sum_i\sum_j \mu_{ij} \qquad \mu_{i\cdot} = \frac{1}{J}\sum_j \mu_{ij} \qquad \mu_{\cdot j} = \frac{1}{I}\sum_i \mu_{ij} \tag{11.7}$$

Por consiguiente μ es la respuesta esperada promedio a todos los niveles de ambos factores (la media grande real) $\mu_{i\cdot}$ es la respuesta esperada promediada a todos los niveles del segundo factor cuando el primer factor A se mantiene en el nivel i y asimismo para $\mu_{\cdot j}$.

DEFINICIÓN

$$\alpha_i = \mu_{i\cdot} - \mu = \textbf{efecto del factor } \boldsymbol{A} \textbf{ al nivel } \boldsymbol{i}$$
$$\beta_j = \mu_{\cdot j} - \mu = \textbf{efecto del factor } \boldsymbol{B} \textbf{ al nivel } \boldsymbol{j}$$
$$\gamma_{ij} = \mu_{ij} - (\mu + \alpha_i + \beta_j) = \textbf{interacción entre el factor } \boldsymbol{A} \textbf{ al nivel } \boldsymbol{i} \textbf{ y el factor } \boldsymbol{B} \textbf{ al nivel } \boldsymbol{j} \tag{11.8}$$

de donde

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} \tag{11.9}$$

El modelo es aditivo si y sólo si todas las $\gamma_{ij} = 0$. Las γ_{ij} se conocen como **parámetros de interacción.** Las α_i se llaman **efectos principales para el factor *A*** y las β_j son los **efectos principales para el factor *B*.** Aunque existen I α_i, J β_i e IJ γ_{ij} además de μ, las condiciones $\sum\alpha_i = 0$, $\sum\beta_j = 0$, $\Sigma_j\gamma_{ij} = 0$ con cualquier i y $\Sigma_i\gamma_{ij} = 0$ cualquier j [todas en virtud de (11.7) y (11.8) implican que sólo IJ de estos nuevos parámetros están independientemente determinados: μ, $I - 1$ de las α_i, $J - 1$ de las β_j e $(I - 1)(J - 1)$ de las γ_{ij}.

Ahora hay tres conjuntos de hipótesis que se considerarán:

H_{0AB}: $\gamma_{ij} = 0$ para todas las i, j	contra	H_{aAB}: por lo menos una $\gamma_{ij} \neq 0$
H_{0A}: $\alpha_1 = \cdots = \alpha_I = 0$	contra	H_{aA}: por lo menos una $\alpha_i \neq 0$
H_{0B}: $\beta_1 = \cdots = \beta_J = 0$	contra	H_{aB}: por lo menos una $\beta_j \neq 0$

Normalmente primero se prueba la hipótesis de no interacción H_{0AB}. Si H_{0AB} no es rechazada, entonces se prueban las otras dos hipótesis para ver si los efectos principales son significativos. Si H_{0AB} es rechazada y H_{0A} se prueba y no es rechazada, el modelo resultante $\mu_{ij} = \mu + \beta_j + \gamma_{ij}$ no se presta para la interpretación directa. En ese caso, es mejor construir un esquema similar a la figura 11.1(b) para tratar de visualizar una forma en la cual interactúan los factores.

El modelo y los procedimientos de prueba

Ahora se utilizan subíndices triples tanto para variables aleatorias como para valores observados, con X_{ijk} y x_{ijk} refiriéndose a la k-ésima observación (replicación) cuando el factor A está al nivel i y el B al factor j.

El modelo de efectos fijos es entonces

$$X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk} \qquad (11.10)$$
$$i = 1, \ldots, I, \quad j = 1, \ldots, J, \quad k = 1, \ldots, K$$

donde las ϵ_{ijk} son independientes y normalmente distribuidas, cada una con media 0 y varianza σ^2.

De nueva cuenta un punto en lugar de un subíndice denota suma de todos los valores del subíndice y una raya horizontal indica promediar. Por consiguiente $X_{ij.}$ es el total de todas las K observaciones realizadas para el factor A al nivel i y el factor B al nivel j [todas las observaciones en la celda (i, j)ésima] y $\overline{X}_{ij.}$ es el promedio de estas K observaciones. Los procedimientos de prueba se basan en las siguientes sumas de cuadrados.

DEFINICIÓN

$$\text{SST} = \sum_i \sum_j \sum_k (X_{ijk} - \overline{X}_{...})^2 \quad \text{grados de libertad} = IJK - 1$$

$$\text{SSE} = \sum_i \sum_j \sum_k (X_{ijk} - \overline{X}_{ij.})^2 \quad \text{grados de libertad} = IJ(K - 1)$$

$$\text{SSA} = \sum_i \sum_j \sum_k (\overline{X}_{i..} - \overline{X}_{...})^2 \quad \text{grados de libertad} = I - 1$$

$$\text{SSB} = \sum_i \sum_j \sum_k (\overline{X}_{.j.} - \overline{X}_{...})^2 \quad \text{grados de libertad} = J - 1$$

$$\text{SSAB} = \sum_i \sum_j \sum_k (\overline{X}_{ij.} - \overline{X}_{i..} - \overline{X}_{.j.} + \overline{X}_{...})^2 \quad \text{grados de libertad} = (I - 1)(J - 1)$$

La identidad fundamental es

$$\text{SST} = \text{SSA} + \text{SSB} + \text{SSAB} + \text{SSE}$$

SSAB se conoce como **interacción de las sumas de cuadrados.**

La variación total se divide por consiguiente en cuatro partes; no explicada (SSE; la cual estaría presente si cualquiera de las tres hipótesis nulas era verdadera o no) y en tres partes que pueden ser explicadas por la verdad o falsedad de las tres H_0. Cada uno de los cuatro cuadrados medios se define como MS = SS/gl. Las medias cuadráticas esperadas sugieren que cada conjunto de hipótesis deberá ser probado utilizando la proporción apropiada de medias cuadráticas con MSE en el denominador:

$$E(\text{MSE}) = \sigma^2$$

$$E(\text{MSA}) = \sigma^2 + \frac{JK}{I-1}\sum_{i=1}^{I}\alpha_i^2 \qquad E(\text{MSB}) = \sigma^2 + \frac{IK}{J-1}\sum_{j=1}^{J}\beta_j^2$$

$$E(\text{MSAB}) = \sigma^2 + \frac{K}{(I-1)(J-1)}\sum_{i=1}^{I}\sum_{j=1}^{J}\gamma_{ij}^2$$

Se puede demostrar que cada una de las tres proporciones de cuadrados medios tiene una distribución F cuando la H_0 asociada es verdadera, lo cual da los siguientes procedimientos de prueba a nivel α.

Hipótesis			Valor estadístico de prueba	Región de rechazo
H_{0A}	contra	H_{aA}	$f_A = \dfrac{\text{MSA}}{\text{MSE}}$	$f_A \geq F_{\alpha,I-1,IJ(K-1)}$
H_{0B}	contra	H_{aB}	$f_B = \dfrac{\text{MSB}}{\text{MSE}}$	$f_B \geq F_{\alpha,J-1,IJ(K-1)}$
H_{0AB}	contra	H_{aAB}	$f_{AB} = \dfrac{\text{MSAB}}{\text{MSE}}$	$f_{AB} \geq F_{\alpha,(I-1)(J-1),IJ(K-1)}$

Ejemplo 11.7 Se ha encontrado que una mezcla de agregado ligero de asfalto tiene una menor conductividad térmica que una mezcla convencional, lo cual es deseable. El artículo "Influence of Selected Mix Design Factors on the Thermal Behavior of Lightweight Aggregate Asphalt Mixes" (*J. of Testing and Eval.*, 2008: 1–8) informó de un experimento en el que se determinaron varias propiedades térmicas de las mezclas. Tres grados diferentes de carpeta fueron usados en combinación con tres diferentes contenidos de agregado grueso (%), con dos observaciones para cada combinación, resultando en los datos de conductividad (W/m· °K) que aparecen en la tabla 11.6.

Tabla 11.6 Datos de conductividad para el ejemplo 11.7

	Contenido de agregado grueso (%)			
Grado de la carpeta de asfalto	**38**	**41**	**44**	$\bar{x}_{i\cdot\cdot}$
PG58	.835, .845	.822, .826	.785, .795	.8180
PG64	.855, .865	.832, .836	.790, .800	.8297
PG70	.815, .825	.800, .820	.770, .790	.8033
$\bar{x}_{\cdot j\cdot}$	.8400	.8227	.7883	

En este caso, $I = J = 3$ y $K = 2$, para un total de $IJK = 18$ observaciones. Los resultados del análisis se resumen en la tabla ANOVA que aparece como tabla 11.7 (una tabla con información adicional aparece en el artículo citado).

Tabla 11.7 Tabla para el ejemplo 11.7

Fuente	Grados de libertad	SS	MS	*f*	*P*
Grado de asfalto	2	.0020893	.0010447	14.12	0.002
Contenido de agregado	2	.0082973	.0041487	56.06	0.000
Interacción	4	.0003253	.0000813	1.10	0.414
Error	9	.0006660	.0000740		
Total	17	.0113780			

El valor de P para la prueba de la presencia de efectos de interacción es .414, que es claramente más grande que cualquier nivel de significacia razonable. Por otra parte, $f_{AB} = 1.10 < F_{.10,4,9} = 2.69$ así que la hipótesis nula de interacción no puede ser rechazada incluso en el nivel de más importancia que sería utilizado en la práctica. Así pues, parece que no hay interacción entre los dos factores. Sin embargo, ambos efectos principales son significativos en el nivel de significancia del 5% (.002 $\leq$ .05 y .000 $\leq$ .05; alternativamente, ambas razones F exceden en gran medida $F_{.05,2,9} = 4.26$). Así que parece que la conductividad promedio real depende de qué grado se utiliza, así como sobre el nivel del contenido de agregado grueso.

La figura 11.5 (a) muestra una gráfica de interacción para los datos de conductividad. Advierta los conjuntos de segmentos de recta casi paralelos para los tres diferentes grados de asfalto, de acuerdo con la prueba F que no muestra efectos de interacción significativos. La conductividad promedio real parece disminuir a medida que disminuye el contenido de agregado. La figura 11.5(b) muestra un diagrama de interacción para la *difusividad térmica* cuyos valores aparecen en el artículo citado. Los dos conjuntos de segmentos de recta de la parte inferior están a punto de ser paralelos, pero difieren marcadamente para PG64; de hecho, la razón F para los efectos de interacción es muy importante aquí.

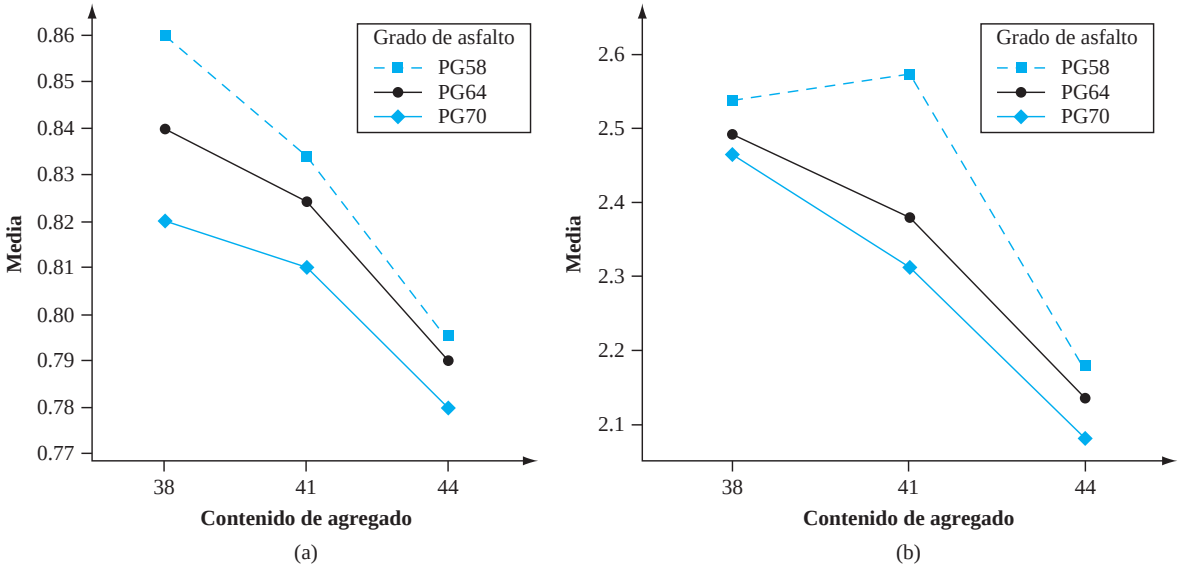


Figura 11.5 Gráficas de interacción para los datos de asfalto del ejemplo 11.7. (a) La variable de respuesta es conductividad. (b) La variable de respuesta es difusividad.

Para verificar la plausibilidad de las suposiciones de normalidad y varianza constante se pueden construir gráficas similares a aquellas de la sección 11.1. Defínanse los valores pronosticados (valores ajustados) como las medias de celdas $\hat{x}_{ijk} = \bar{x}_{ij.}$. Por ejemplo el valor predicho para el grado PG58 y un contenido de agregado 38 es $\hat{x}_{11k} = (.835 + .845)/2 = .840$ para $k = 1, 2$. Los residuos son las diferencias entre las observaciones y los valores pronosticados correspondientes: $x_{ijk} - \bar{x}_{ij.}$. La gráfica de probabilidad

normal de los residuos se muestra en la figura 11.6(a). El patrón es suficientemente lineal por lo que no hay que preocuparse por la suposición de normalidad. La gráfica de residuos contra los valores ajustados en la figura 11.6(b) muestra una dispersión un poco menor en el lado derecho que en el izquierdo pero no suficiente para ser una diferencial preocupante, por lo que una varianza constante es una suposición razonable.

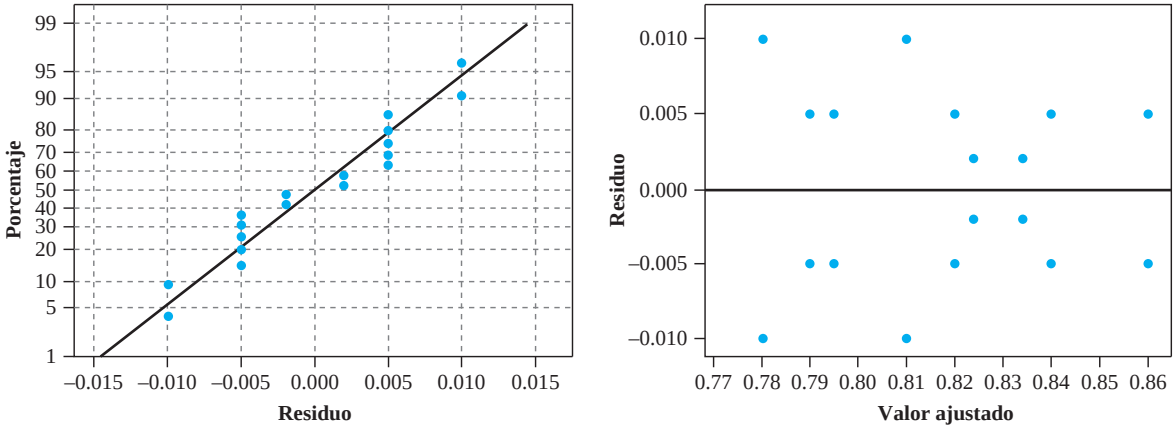


Figura 11.6 Gráficas para verificar las suposiciones de normalidad y varianza constante para el ejemplo 11.7 ■

Comparaciones múltiples

Cuando la hipótesis de no interacción H_{0AB} no es rechazada y por lo menos una de las dos hipótesis nulas de efecto principal es rechazada, se puede utilizar el método de Tukey para identificar diferencias significativas en los niveles. Para identificar diferencias entre las α_i cuando H_{0A} es rechazada,

1. Se obtiene $Q_{\alpha,I,IJ(K-1)}$, donde el segundo subíndice J identifica el número de niveles que se están comparando y el tercero se refiere al número de grados de libertad en cuanto al error.
2. Se calcula $w = Q\sqrt{\text{MSE}/(JK)}$, donde JK es el número de observaciones promediadas para obtener cada una de las $\bar{x}_{i\cdot\cdot}$ en el paso 3.
3. Se ordenan las $\bar{x}_{i\cdot\cdot}$ desde la más pequeña hasta la más grande, se subrayan todos los pares que difieren por menos de w. Los pares no subrayados corresponden a niveles del factor A significativamente diferentes.

Para identificar niveles diferentes del factor B cuando H_{0B} es rechazada, se reemplaza el segundo subíndice en Q por J, se reemplaza JK por IK en w y se reemplaza $\bar{x}_{i\cdot\cdot}$ por $\bar{x}_{\cdot j\cdot}$.

Ejemplo 11.8 (Continuación del ejemplo 11.7)

$I = J = 3$, tanto para el factor A (grado) y el factor B (contenido de agregado). Con $\alpha = 0.05$ y gl error $= IJ(K-1) = 9$, $Q_{.05,3,9} = 3.95$. El criterio para la identificación de diferencias significativas es entonces $w = 3.95\sqrt{.0000740/6} = .00139$. El *grado* de la media muestral en orden creciente es .8033, .8180 y .8297. Sólo la diferencia entre las dos principales medias es menor que w. Esto da la pauta para subrayar

PG70 PG58 PG64

Los grados PG58 y PG64 no parecen diferir significativamente uno de otro en el efecto sobre la conductividad promedio real, pero ambos difieren de la calificación PG70.

Las medias ordenadas para el factor B son .7883, .8227 y .8400. Los tres pares de medias difieren en más de .00139, por lo que no hay líneas de subrayado. La conductividad promedio real parece ser diferente para los tres niveles de contenido agregado. ■

Modelos con efectos combinados y aleatorios

En algunos problemas, es posible que los niveles de uno u otro factor hayan sido seleccionados de una gran población de niveles posibles, así que los efectos contribuidos por el factor son aleatorios en lugar de fijos. Como en la sección 11.1, si ambos factores contribuyen con efectos aleatorios, el modelo se conoce como modelo de efectos aleatorios, en tanto que si un factor es fijo y el otro es aleatorio, resulta un modelo de efectos combinados. Ahora se considerará el análisis de un modelo de efectos combinados en el cual el factor *A* (renglones) es el factor fijo y el factor *B* (columnas) es el factor aleatorio. El caso en el cual ambos factores son aleatorios se aborda en el ejercicio 26.

DEFINICIÓN

El **modelo de efectos combinados** cuando el factor *A* es fijo y el *B* es aleatorio es

$$X_{ijk} = \mu + \alpha_i + B_j + G_{ij} + \epsilon_{ijk}$$

$$i = 1, \ldots, I, \quad j = 1, \ldots, J, \quad k = 1, \ldots, K$$

Aquí μ y α_i son constantes con $\sum\alpha_i = 0$ y las B_j, G_{ij} y ϵ_{ijk} son variables aleatorias independientes normalmente distribuidas con valor esperado 0 y varianzas σ_B^2, σ_G^2 y σ^2, respectivamente.* Las hipótesis relevantes en este caso son algo diferentes de las del modelo de efectos fijos.

H_{0A}: $\alpha_1 = \alpha_2 = \cdots = \alpha_I = 0$	contra	H_{aA}: por lo menos un $\alpha_i \neq 0$
H_{0B}: $\sigma_B^2 = 0$	contra	H_{aB}: $\sigma_B^2 > 0$
H_{0G}: $\sigma_G^2 = 0$	contra	H_{aG}: $\sigma_G^2 > 0$

Se acostumbra probar H_{0A} y H_{0B} sólo si la hipótesis de no interacción H_{0G} no puede ser rechazada.

Las sumas de cuadrados y cuadrados medios requeridas para los procedimientos de prueba se definen y calculan con exactitud como en el caso de efectos fijos. Los cuadrados medios esperados para el modelo combinado son

$$E(\text{MSE}) = \sigma^2$$

$$E(\text{MSA}) = \sigma^2 + K\sigma_G^2 + \frac{JK}{I-1}\sum\alpha_i^2$$

$$E(\text{MSB}) = \sigma^2 + K\sigma_G^2 + IK\sigma_B^2$$

$$E(\text{MSAB}) = \sigma^2 + K\sigma_G^2$$

La razón $f_{AB} = \text{MSAB/MSE}$ es de nuevo aplicable para probar la hipótesis de no interacción, con H_{0G} rechazada si $f_{AB} \geq F_{\alpha,(I-1)(J-1),IJ(K-1)}$. Sin embargo, para probar H_{0A} contra H_{aA}, los cuadrados medios esperados sugieren que aunque el numerador de la razón *F* deberá seguir siendo MSA, el denominador deberá ser MSAB en lugar de MSE. MSAB también es el denominador de la razón *F* para probar H_{0B}.

* A esto se le llama modelo "no restringido". Un modelo "restringido" alternativo requiere que $\Sigma_i G_{ij} = 0$ para toda *j* (así las G_{ij} dejan de ser independientes). Las medias cuadráticas esperadas y las razones *F* apropiadas para probar ciertas hipótesis dependen del modelo seleccionado. La opción por defecto de Minitab da respuesta al modelo no restringido.

Para probar H_{0A} contra H_{aA} (los factores A fijo, B aleatorio), el valor del estadístico de prueba es $f_A = \text{MSA/MSAB}$ y la región de rechazo es $f_A \geq F_{\alpha,I-1,(I-1)(J-1)}$. La prueba de H_{0B} contra H_{aB} utiliza $f_B = \text{MSB/MSAB}$ y la región de rechazo es $f_B \geq F_{\alpha,J-1,(I-1)(J-1)}$.

Ejemplo 11.9 Un ingeniero de procesos ha identificado dos causas potenciales de vibración de motores eléctricos, el material utilizado para la carcasa del motor (factor A) y el proveedor de cojinetes utilizados en el motor (factor B). Los datos adjuntos sobre la cantidad de vibración (micrones) se obtuvieron con un experimento en el cual se construyeron motores con carcasas de acero, aluminio y plástico y cojinetes suministrados por cinco proveedores seleccionados al azar.

		Proveedor				
		1	**2**	**3**	**4**	**5**
Material	**Acero**	13.1 13.2	16.3 15.8	13.7 14.3	15.7 15.8	13.5 12.5
	Aluminio	15.0 14.8	15.7 16.4	13.9 14.3	13.7 14.2	13.4 13.8
	Plástico	14.0 14.3	17.2 16.7	12.4 12.3	14.4 13.9	13.2 13.1

Sólo los tres materiales para carcasas utilizados en el experimento se están considerando para usarse en la producción, de ahí que el factor A es fijo. Sin embargo, los cinco proveedores fueron seleccionados al azar de una población mucho más grande, de modo que el factor B es aleatorio. Las hipótesis nulas relevantes son

$$H_{0A}: \alpha_1 = \alpha_2 = \alpha_3 = 0 \qquad H_{0B}: \sigma_B^2 = 0 \qquad H_{0AB}: \sigma_G^2 = 0$$

En la figura 11.7 aparecen los resultados generados por Minitab. La columna de valor P en la tabla ANOVA indica que las últimas dos hipótesis nulas deberán ser rechazadas a un nivel de significancia de .05. Los diferentes materiales para carcasas por sí mismos no parecen afectar la vibración, pero la interacción entre el material y el proveedor es una causa significativa de la variación de la vibración.

```
Factor     Tipo      Niveles    Valores
Matcar     fijo            3          1    2    3
fuente aleatorio           5          1    2    3    4    5

Fuente               GL          SS          MS          F          P
Matcar                2      0.7047      0.3523       0.24      0.790
Fuente                4     36.6747      9.1687       6.32      0.013
Matcar*fuente         8     11.6053      1.4507      13.03      0.000
Error                15      1.6700      0.1113
Total                29     50.6547

Fuente             Componente   Término   Media cuadrática esperada para cada término
                  de varianza   de error  (usando modelo no restringido)
1 matcar                           3      (4)+2(3)+Q[1]
2 fuente                1.2863     3      (4)+2(3)+6(2)
3 matcar*fuente         0.6697     4      (4)+2(3)
4 Error                 0.1113            (4)
```

Figura 11.7 Salida de Minitab para la opción ANOVA balanceado para los datos del ejemplo 11.9 ■

Cuando por lo menos dos de las K_{ij} son desiguales, los cálculos ANOVA son mucho más complejos que en el caso $K_{ij} = K$, y existe controversia en cuanto a cuáles procedimientos de prueba deben ser utilizados. Una de las referencias del capítulo puede ser consultada para más información.

EJERCICIOS Sección 11.2 (16–26)

16. En un experimento para valuar los efectos de tiempo de fraguado (factor *A*) y tipo de mezcla (factor *B*) en la resistencia a la compresión de cubos de cemento endurecido, se utilizaron tres tiempos de fraguado diferentes en combinación con cuatro mezclas diferentes, con tres observaciones obtenidas para cada una de las 12 combinaciones de mezcla–tiempo de fraguado. Las sumas de cuadrados resultantes fueron SSA = 30,763.0, SSB = 34,185.6, SSE = 97,436.8 y SST = 205,966.6.

a. Construya una tabla ANOVA.

b. Pruebe al nivel .05 la hipótesis nula H_{0AB}: todas las $\gamma_{ij} = 0$ (ninguna interacción de factores) contra H_{aAB}: por lo menos una $\gamma_{ij} \neq 0$.

c. Pruebe al nivel .05 la hipótesis nula H_{0AB}: $\alpha_1 = \alpha_2 = \alpha_3 = 0$ (efectos principales del factor *A* ausentes) contra H_{aA} por lo menos una $\alpha_i \neq 0$.

d. Pruebe H_{0B}: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ contra H_{aB}: por lo menos una $\beta_j \neq 0$ utilizando una prueba al nivel .05.

e. Los valores de las $\bar{x}_{i..}$ fueron $\bar{x}_{1..} = 4010.88$, $\bar{x}_{2..} = 4029.10$ y $\bar{x}_{3..} = 3960.02$. Use el procedimiento de Tukey para investigar diferencias significativas entre los tres tiempos de fraguado.

17. El artículo "Towards Improving the Properties of Plaster Moulds and Castings" (*J. Engr. Manuf.*, 1991: 265–269) describe varios ANOVAs realizados para estudiar cómo la cantidad de fibra de carbón y las adiciones de arena afectan las diversas características del proceso de moldeo. A continuación se dan datos sobre dureza de la pieza moldeada y resistencia del molde húmedo.

Adición de arena (%)	Adición de fibra de carbón (%)	Dureza de la pieza fundida	Resistencia de molde húmedo
0	0	61.0	34.0
0	0	63.0	16.0
15	0	67.0	36.0
15	0	69.0	19.0
30	0	65.0	28.0
30	0	74.0	17.0
0	.25	69.0	49.0
0	.25	69.0	48.0
15	.25	69.0	43.0
15	.25	74.0	29.0
30	.25	74.0	31.0
30	.25	72.0	24.0
0	.50	67.0	55.0
0	.50	69.0	60.0
15	.50	69.0	45.0
15	.50	74.0	43.0
30	.50	74.0	22.0
30	.50	74.0	48.0

a. Un ANOVA para la resistencia del molde húmedo da SSArena = 705, SSFibra = 1278, SSE = 843 y SST = 3105. Pruebe en busca de cualesquiera otros efectos con $\alpha = .05$.

b. Realice un ANOVA de las observaciones de dureza de la pieza fundida con $\alpha = .05$.

c. Grafique la dureza media muestral contra el porcentaje de arena con niveles diferentes de fibra de carbón. ¿Es la gráfica consistente con su análisis en el inciso (b)?

18. Los datos adjuntos se obtuvieron con un experimento para investigar si el rendimiento con un cierto proceso químico dependía de la formulación de una entrada particular o de la velocidad del mezclador.

		Velocidad		
		60	**70**	**80**
Formulación	**1**	189.7 188.6 190.1	185.1 179.4 177.3	189.0 193.0 191.1
	2	165.1 165.9 167.6	161.7 159.8 161.6	163.3 166.6 170.3

Un programa estadístico dio SS(Form) = 2253.44, SS(Velocidad) = 230.81, SS(Forma*Velocidad) = 18.58 y SSE = 71.87.

a. ¿Parece haber interacción entre los factores?

b. ¿Parece que el rendimiento depende de la formulación o la velocidad?

c. Calcule las estimaciones de los efectos principales.

d. Los *valores ajustados* son $\hat{x}_{ijk} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_{ij}$ y los *residuos* son $x_{ijk} - \hat{x}_{ijk}$. Verifique que los residuos sean .23, −.87, .63, 4.50, −1.20, −3.30, −2.03, 1.97, .07, −1.10, −.30, 1.40, .67, −1.23, .57, −3.43, −.13, y 3.57.

e. Construya una gráfica de probabilidad normal con los residuos dados en el inciso (d). ¿Parecen estar normalmente distribuidas las ϵ_{ijk}?

19. La tabla de datos adjuntos da observaciones sobre acidez total de muestras de carbón de tres tipos diferentes, con las determinaciones obtenidas con tres concentraciones diferentes de NaOH etanólico ("Chemistry of Brown Coals", *Australian J. Applied Science*, 1958: 375–379).

		Tipo de carbón		
		Morwell	**Yallourn**	**Maddingley**
Concentración de NaOH	**.404N**	8.27, 8.17	8.66, 8.61	8.14, 7.96
	.626N	8.03, 8.21	8.42, 8.58	8.02, 7.89
	.786N	8.60, 8.20	8.61, 8.76	8.13, 8.07

a. Suponiendo que ambos efectos son fijos, construya una tabla ANOVA, pruebe en cuanto a la presencia de interacción y luego en cuanto a la presencia de efectos principales por cada factor (todo a un nivel de .01).

b. Use un procedimiento de Tukey para identificar diferencias significativas entre los tipos de carbón.

20. El artículo "Fatigue Limits of Enamel Bonds with Moist and Dry Techniques" (*Dental Materials*, 2009: 1527–1531) describe un experimento para investigar la capacidad de los sistemas adhesivos para unir las estructuras mineralizadas del diente. La respuesta variable es la fuerza de cizalla (MPa) y dos diferentes adhesivos (Adper Single Bond Plus y OptiBond Solo Plus) se utilizan en combinación con dos diferentes condiciones de la superficie. Los datos adjuntos fueron suministrados por los autores del artículo. Las primeras 12 observaciones vinieron del tratamiento SBP-seco, los próximos 12 del tratamiento PAS-húmedo, los siguientes 12 tratamiento de OBP-seco y los últimos 12 proceden del tratamiento OBP-húmedo.

56.7	57.4	53.4	54.0	49.9	49.9
56.2	51.9	49.6	45.7	56.8	54.1
49.2	47.4	53.7	50.6	62.7	48.8
41.0	57.4	51.4	53.4	55.2	38.9
38.8	46.0	38.0	47.0	46.2	39.8
25.9	37.8	43.4	40.2	35.4	40.3
40.6	35.5	58.7	50.4	43.1	61.7
33.3	38.7	45.4	47.2	53.3	44.9

a. Construya un diagrama de caja comparativo de los datos de los cuatro tratamientos diferentes y coméntelo.

b. Lleve a cabo un análisis adecuado de la varianza y exprese sus conclusiones (utilice un nivel de significancia de .01 para las pruebas). Incluya todos los gráficos que permitan apreciarlo.

c. Si un nivel de significancia de .05 se utiliza para el ANOVA de dos vías, el efecto de la interacción es significativo (al igual que en general pegamentos diferentes funcionan mejor con unos materiales que con otros). Así que ahora tiene sentido llevar a cabo un ANOVA unidireccional en los cuatro tratamientos PAS-D, PAS-M, OBP-D y OBP-M. Haga esto y determine las diferencias significativas entre los tratamientos.

21. En un experimento para investigar el efecto del "factor cemento" (número de sacos de cemento por yarda cúbica) en la resistencia a la flexión del concreto resultante ("Studies of Flexural Strength of Concrete, Part 3: Effects of Variation in Testing Procedure", *Proceedings ASTM*, 1957: 1127–1139), se utilizaron $I = 3$ valores de factor diferentes, se seleccionaron $J = 5$ lotes diferentes de cemento y se vaciaron $K = 2$ vigas con cada combinación de factor cemento/lote. Las sumas de cuadrados incluyen SSA = 22,941.80, SSB = 22,765.53, SSE = 15,253.50 y SST = 64,954.70. Construya la tabla ANOVA. Entonces suponiendo un modelo combinado con factor de cemento (A) fijo y lotes (B) aleatorios, pruebe los tres pares de hipótesis de interés al nivel .05.

22. Se realizó un estudio para comparar las vidas útiles de escritura de cuatro marcas premium de plumas. Se pensaba que la superficie de escritura podría afectar la vida útil, así que se seleccionaron al azar tres superficies diferentes. Se utilizó una máquina de escritura para asegurar que las condiciones permanecieran homogéneas (p. ej., presión constante y un ángulo fijo). La tabla adjunta muestra las dos vidas útiles (min) obtenidas con cada combinación de marca–superficie.

Superficie de escritura

		1	2	3	$x_{i..}$
Marca de pluma	**1**	709, 659	713, 726	660, 645	4112
	2	668, 685	722, 740	692, 720	4227
	3	659, 685	666, 684	678, 750	4122
	4	698, 650	704, 666	686, 733	4137
	$x_{.j.}$	5413	5621	5564	16,598

Realice un ANOVA apropiado y exprese su conclusión.

23. Los datos adjuntos se obtuvieron en un experimento para investigar si la resistencia a la compresión de cilindros de concreto depende del tipo de material de remate utilizado o de la variabilidad de los diferentes lotes ("The Effect of Type of Capping Material on the Compressive Strength of Concrete Cilinders", *Proceedings ASTM*, 1958: 1166–1186). Cada número es un total de celda ($x_{ij.}$) basado en $K = 3$ observaciones.

Lote

		1	2	3	4	5
Material de remate	**1**	1847	1942	1935	1891	1795
	2	1779	1850	1795	1785	1626
	3	1806	1892	1889	1891	1756

Además, $\sum\sum\sum x_{ijk}^2 = 16{,}815{,}853$ y $\sum\sum x_{ij.}^2 = 50{,}443{,}409$. Obtenga la tabla ANOVA y luego pruebe al nivel .01 las hipótesis H_{0G} contra H_{aG}, H_{0A} contra H_{aA} y H_{0B} contra H_{aB}, suponiendo que el remate es un efecto fijo y los lotes son un efecto aleatorio.

24. **a.** Demuestre que $E(\overline{X}_{i..} - \overline{X}_{...}) = \alpha_i$ de modo que $\overline{X}_{i..} - \overline{X}_{...}$ es un estimador insesgado de α_i (en el modelo de efectos fijos.

b. Con $\hat{\gamma}_{ij} = \overline{X}_{ij.} - \overline{X}_{i..} - \overline{X}_{.j.} + \overline{X}_{...}$, demuestre que $\hat{\gamma}_{ij}$ es un estimador insesgado de γ_{ij} (en el modelo de efectos fijos).

25. Demuestre cómo se puede obtener un intervalo de confianza t de $100(1 - \alpha)\%$ para $\alpha_i - \alpha_{i'}$. Luego calcule un intervalo de 95% para $\alpha_2 - \alpha_3$ con los datos del ejercicio 19. [*Sugerencia*: con $\theta = \alpha_2 - \alpha_3$, el resultado del ejercicio 24(a) indica cómo obtener $\hat{\theta}$. En seguida calcule $V(\hat{\theta})$ y $\sigma_{\hat{\theta}}$, y obtenga una estimación de $\sigma_{\hat{\theta}}$ con $\sqrt{\text{MSE}}$ para estimar σ (la cual identifica el número de grados de libertad apropiado).]

26. Cuando ambos factores son aleatorios en un experimento ANOVA bidireccional con K réplicas por cada combinación de niveles de factor, los cuadrados medios esperados $E(\text{MSE}) = \sigma^2$, $E(\text{MSA}) = \sigma^2 + K\sigma_G^2 + JK\sigma_A^2$, $E(\text{MSB}) = \sigma^2 + K\sigma_G^2 + IK\sigma_B^2$ y $E(\text{MSAB}) = \sigma^2 + K\sigma_G^2$.

a. ¿Qué razón F es apropiada para probar H_{0G}: $\sigma_G^2 = 0$ contra H_{aG}: $\sigma_G^2 > 0$?

b. Responda el inciso (a) para probar H_{0A}: $\sigma_A^2 = 0$ contra H_{aA}: $\sigma_A^2 > 0$ y H_{0B}: $\sigma_B^2 = 0$ contra H_{aB}: $\sigma_G^2 > 0$.

11.3 ANOVA con tres factores

Para indicar la naturaleza de los modelos y análisis cuando los experimentos ANOVA implican más de dos factores, aquí se abordará el caso de tres factores fijos–*A*, *B*, y *C*. Los números de niveles de los tres factores se denotarán por *I*, *J* y *K*, respectivamente y L_{ijk} = el número de observaciones realizadas con el factor *A* al nivel *i*, el factor *B* al nivel *j* y el factor *C* al nivel *k*. El análisis es bastante complicado cuando las L_{ijk} no son iguales, por lo que se hace $L_{ijk} = L$. En ese caso X_{ijkl} y x_{ijkl} denotan el valor observado, antes y después de que se realiza el experimento de la *l*-ésima réplica ($l = 1, 2, \ldots, L$) cuando los tres factores están fijos en los niveles *i*, *j* y *k*.

Para entender los parámetros que aparecerán en el modelo ANOVA trifactorial, primero recuérdese que en el ANOVA bifactorial con réplicas, $E(X_{ijk}) = \mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$, donde las restricciones $\Sigma_i\alpha_i = \Sigma_j\beta_j = 0$, $\Sigma_i\gamma_{ij} = 0$ por cada *j* y $\Sigma_j\gamma_{ij} = 0$ por cada *i* eran necesarias para obtener un conjunto único de parámetros. Si se utilizan subíndices puntuales en las μ_{ij} para denotar el cálculo de promedios (en lugar de suma), entonces

$$\mu_{i\cdot} - \mu_{\cdot\cdot} = \frac{1}{J}\sum_j \mu_{ij} - \frac{1}{IJ}\sum_i\sum_j \mu_{ij} = \alpha_i$$

es el efecto del factor *A* al nivel *i* promediado a todos los niveles del factor *B*, mientras que

$$\mu_{ij} - \mu_{\cdot j} = \mu_{ij} - \frac{1}{I}\sum_i \mu_{ij} = \alpha_i + \gamma_{ij}$$

es el efecto del factor *A* al nivel *i* específico del factor *B* al nivel *j*. Si el efecto de *A* al nivel *i* depende del nivel de *B*, entonces existe interacción entre los factores y las γ_{ij} no son todas cero. En particular,

$$\mu_{ij} - \mu_{\cdot j} - \mu_{i\cdot} + \mu_{\cdot\cdot} = \gamma_{ij} \tag{11.11}$$

Modelo de efectos fijos y procedimientos de prueba

El **modelo de efectos fijos para ANOVA con tres factores** con $L_{ijk} = L$ es

$$X_{ijkl} = \mu_{ijk} + \epsilon_{ijkl} \quad \begin{matrix} i = 1, \ldots, I, & j = 1, \ldots, J \\ k = 1, \ldots, K, & l = 1, \ldots, L \end{matrix} \tag{11.12}$$

donde las ϵ_{ijkl} están normalmente distribuidas con media 0 y varianza σ^2 y

$$\mu_{ijk} = \mu + \alpha_i + \beta_i + \delta_k + \gamma_{ij}^{AB} + \gamma_{ik}^{AC} + \gamma_{jk}^{BC} + \gamma_{ijk} \tag{11.13}$$

Las restricciones necesarias para obtener parámetros unívocamente definidos son que la suma para cualquier subíndice de cualquier parámetro a la derecha de (11.13) sea igual a cero.

Los parámetros γ_{ij}^{AB}, γ_{ik}^{AC} y γ_{jk}^{BC} se llaman interacciones de dos factores y γ_{ijk} se llama interacción de tres factores; las α_i, β_j y δ_k son los parámetros de los efectos principales. A cualquier nivel fijo *k* del tercer factor, análogo a (11.11),

$$\mu_{ijk} - \mu_{i\cdot k} - \mu_{\cdot jk} + \mu_{\cdot\cdot k} = \gamma_{ij}^{AB} + \gamma_{ijk}$$

es la interacción del *i*-ésimo nivel de *A* con el *j*-ésimo nivel de *B* específico del *k*-ésimo nivel de *C*, mientras que

$$\mu_{ij\cdot} - \mu_{i\cdot\cdot} - \mu_{\cdot j\cdot} + \mu_{\cdots} = \gamma_{ij}^{AB}$$

es la interacción entre A al nivel i y B al nivel j promediada a todos los niveles de C. Si la interacción de A al nivel i y B al nivel j no depende de k, entonces todas las γ_{ijk} son iguales a 0. Por tanto las γ_{ijk} diferentes de cero representan no aditividad de las γ_{ij}^{AB} para dos factores a los varios niveles del tercer factor C. Si el experimento incluyó más de tres factores, habría términos de interacción de mayor grado correspondientes con interpretaciones análogas. Obsérvese que en el argumento previo, si se hubiera considerado fijar el nivel de A o B (en lugar del de C, como se hizo) y examinando las γ_{ijk} su interpretación sería la misma; si cualquiera de las interacciones de dos factores dependen del nivel del tercer factor, entonces hay γ_{ijk} no nulas.

Cuando $L > 1$, existe una suma de cuadrados por cada efecto principal, por cada interacción de dos factores y por la interacción de tres factores. Para escribir éstas en una forma que indique cómo se definen las sumas de cuadrados cuando existen más de tres factores, obsérvese que cualquiera de los parámetros de modelo en (11.13) puede ser estimado insesgadamente promediando X_{ijkl} para los subíndices apropiados y considerando las diferencias. Por lo tanto

$$\hat{\mu} = \overline{X}_{....} \qquad \hat{\alpha}_i = \overline{X}_{i...} - \overline{X}_{....} \qquad \hat{\gamma}_{ij}^{AB} = \overline{X}_{ij..} - \overline{X}_{i...} - \overline{X}_{.j..} + \overline{X}_{....}$$
$$\hat{\gamma}_{ijk} = \overline{X}_{ijk.} - \overline{X}_{ij..} - \overline{X}_{i.k.} - \overline{X}_{.jk.} + \overline{X}_{i...} + \overline{X}_{.j..} + \overline{X}_{..k.} - \overline{X}_{....}$$

con los demás efectos principales y estimadores de interacción obtenidos por simetría.

DEFINICIÓN

Las **sumas de cuadrados** apropiadas son,

$$\text{SST} = \sum_i \sum_j \sum_k \sum_l (X_{ijkl} - \overline{X}_{....})^2 \qquad \text{grados de libertad} = IJKL - 1$$

$$\text{SSA} = \sum_i \sum_j \sum_k \sum_l \hat{\alpha}_i^2 = JKL \sum_i (\overline{X}_{i...} - \overline{X}_{....})^2 \qquad \text{grados de libertad} = I - 1$$

$$\text{SSAB} = \sum_i \sum_j \sum_k \sum_l (\hat{\gamma}_{ij}^{AB})^2 \qquad \text{grados de libertad} = (I-1)(J-1)$$
$$= KL \sum_i \sum_j (\overline{X}_{ij..} - \overline{X}_{i...} - \overline{X}_{j...} + \overline{X}_{....})^2$$

$$\text{SSABC} = \sum_i \sum_j \sum_k \sum_l \hat{\gamma}_{ijk}^2 = L \sum_i \sum_j \sum_k \hat{\gamma}_{ijk}^2 \qquad \text{grados de libertad} = (I-1)(J-1)(K-1)$$

$$\text{SSE} = \sum_i \sum_j \sum_k \sum_l (X_{ijkl} - \overline{X}_{ijk.})^2 \qquad \text{grados de libertad} = IJK(L-1)$$

con los demás efectos principales y las sumas de cuadrados de interacción de dos factores obtenidos por simetría. SST es la suma de las otras ocho sumas de cuadrados (SS).

Cada suma de cuadrados (excepto SST) cuando se divide entre sus grados de libertad da un cuadrado medio. Las medias cuadráticas esperadas son

$$E(\text{MSE}) = \sigma^2$$

$$E(\text{MSA}) = \sigma^2 + \frac{JKL}{I-1} \sum_i \alpha_i^2$$

$$E(\text{MSAB}) = \sigma^2 + \frac{KL}{(I-1)(J-1)} \sum_i \sum_j (\gamma_{ij}^{AB})^2$$

$$E(\text{MSABC}) = \sigma^2 + \frac{L}{(I-1)(J-1)(K-1)} \sum_i \sum_j \sum_k (\gamma_{ijk})^2$$

con expresiones similares de los demás cuadrados medios esperados. El efecto principal y las hipótesis de interacción se prueban formando razones F con MSE en cada denominador.

Hipótesis nula	Valor estadístico de prueba	Región de rechazo
H_{0A}: todas las $\alpha_i = 0$	$f_A = \dfrac{\text{MSA}}{\text{MSE}}$	$f_A \geq F_{\alpha,I-1,IJK(L-1)}$
H_{0AB}: todas las $\gamma_{ij}^{AB} = 0$	$f_{AB} = \dfrac{\text{MSAB}}{\text{MSE}}$	$f_{AB} \geq F_{\alpha,(I-1)(J-1),IJK(L-1)}$
H_{0ABC}: todas las $\gamma_{ijk} = 0$	$f_{ABC} = \dfrac{\text{MSABC}}{\text{MSE}}$	$f_{ABC} \geq F_{\alpha,(I-1)(J-1)(K-1),IJK(L-1)}$

Normalmente las hipótesis de efecto principal se prueban sólo si todas las interacciones no son significativas.

Este análisis asume que $L_{ijk} = L > 1$. Si $L = 1$, entonces como en el caso de dos factores, las interacciones de alto grado deben ser supuestas iguales a 0 para obtener un MSE que estime σ^2. Con $L = 1$ y haciendo caso omiso de la suma para el cuarto subíndice l, las fórmulas anteriores para las sumas de cuadrados continúan siendo válidas y la suma de cuadrados en cuanto a error es $\text{SSE} = \Sigma_i\Sigma_j\Sigma_k\hat{\gamma}_{ijk}^2$ con $\overline{X}_{ijk\cdot} = X_{ijk}$ en la expresión para $\hat{\gamma}_{ijk}$.

Ejemplo 11.10 Las siguientes observaciones (temperatura corporal − 100°F) se reportaron en un experimento para estudiar la tolerancia al calor de ganado ("The Significance of the Coat in Heat Tolerance of Cattle", *Australian J. Agric. Res*, 1959: 744–748). Se realizaron mediciones en cuatro periodos diferentes (factor A, con $I = 4$) en dos razas diferentes de ganado (factor B, con $J = 2$) que tienen cuatro tipos diferentes de pelaje (factor C, con $K = 4$); $L = 3$ observaciones fueron realizadas por cada una de las $4 \times 2 \times 4 = 32$ combinaciones de niveles de los tres factores.

	B_1				B_2			
	C_1	C_2	C_3	C_4	C_1	C_2	C_3	C_4
A_1	3.6	3.4	2.9	2.5	4.2	4.4	3.6	3.0
	3.8	3.7	2.8	2.4	4.0	3.9	3.7	2.8
	3.9	3.9	2.7	2.2	3.9	4.2	3.4	2.9
A_2	3.8	3.8	2.9	2.4	4.4	4.2	3.8	2.0
	3.6	3.9	2.9	2.2	4.4	4.3	3.7	2.9
	4.0	3.9	2.8	2.2	4.6	4.7	3.4	2.8
A_3	3.7	3.8	2.9	2.1	4.2	4.0	4.0	2.0
	3.9	4.0	2.7	2.0	4.4	4.6	3.8	2.4
	4.2	3.9	2.8	1.8	4.5	4.5	3.3	2.0
A_4	3.6	3.6	2.6	2.0	4.0	4.0	3.8	2.0
	3.5	3.7	2.9	2.0	4.1	4.4	3.7	2.2
	3.8	3.9	2.9	1.9	4.2	4.2	3.5	2.3

La tabla de totales de celda ($x_{ijk\cdot}$) con todas las combinaciones de los tres factores es

	B_1				B_2			
$x_{ijk\cdot}$	C1	C2	C3	C4	C1	C2	C3	C4
A_1	11.3	11.0	8.4	7.1	12.1	12.5	10.7	8.7
A_2	11.4	11.6	8.6	6.8	13.4	13.2	10.9	7.7
A_3	11.8	11.7	8.4	5.9	13.1	13.1	11.1	6.4
A_4	10.9	11.2	8.4	5.9	12.3	12.6	11.0	6.5

La figura 11.8 muestra gráficas de las medias de celda correspondientes $\bar{x}_{ijk\cdot} = x_{ijk\cdot}/3$. Se regresará a estas gráficas después de considerar pruebas de varias hipótesis. La base para estas pruebas es la tabla ANOVA dada en la tabla 11.8.

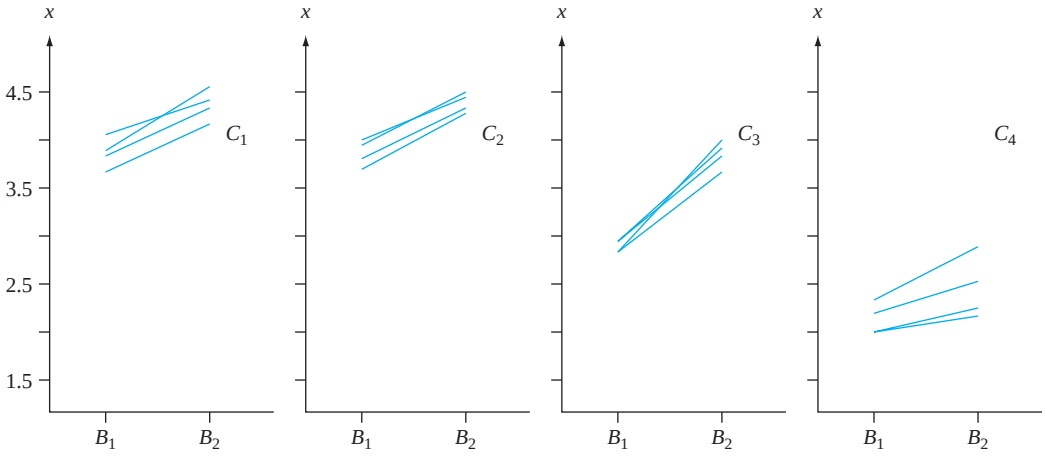


Figura 11.8 Gráficas de $\bar{x}_{ijk\cdot}$ para el ejemplo 11.10

Tabla 11.8 Tabla ANOVA para el ejemplo 11.10

Fuente	gl	Suma de cuadrados	Media cuadrática	f
A	$I - 1 = 3$	.49	.163	4.13
B	$J - 1 = 1$	6.45	6.45	163.29
C	$K - 1 = 3$	48.93	16.31	412.91
AB	$(I - 1)(J - 1) = 3$	.02	.0067	.170
AC	$(I - 1)(K - 1) = 9$	1.61	.179	4.53
BC	$(J - 1)(K - 1) = 3$	.88	.293	7.42
ABC	$(I - 1)(J - 1)(K - 1) = 9$	.25	.0278	.704
Error	$IJK(L - 1) = 64$	2.53	.0395	
Total	$IJKL - 1 = 95$	61.16		

Como $F_{.01,9,64} \approx 2.70$ y $f_{ABC} = \text{MSABC/MSE} = .704$ no excede de 2.70, se concluye que las interacciones de tres factores no son significativas. Sin embargo, aunque también las interacciones AB no son significativas, tanto las interacciones AC como BC así como también los efectos principales parecen ser necesarios en el modelo. Cuando no existen interacciones ABC o AB, una gráfica de las $\bar{x}_{ijk\cdot}(= \hat{\mu}_{ijk})$ por separado para cada nivel de C no deberá revelar ningunas interacciones sustanciales (si sólo las interacciones ABC son cero, las gráficas son más difíciles de interpretar: véase el artículo "Two-Dimensional Plots for Interpreting Interactions in the Three-Factor Analysis of Variance Model", *Amer. Statistician*, mayo de 1979: 63-69). ■

Se pueden construir gráficas de diagnóstico para verificar las suposiciones de normalidad y varianza constante como se describió en secciones previas. Se puede utilizar el procedimiento de Tukey en ANOVA de tres factores (o más). El segundo subíndice en Q es el número de medias muestrales que se están comparando y el tercero es grados de libertad para error.

Los modelos con efectos aleatorios y combinados en ocasiones también son apropiados. Las sumas de cuadrados y grados de libertad son idénticos al caso de efectos fijos, pero los cuadrados medios esperados son, desde luego, diferentes para los efectos principales aleatorios o interacciones. Una buena referencia es el libro de Douglas Montgomery que aparece en la bibliografía del capítulo.

Diseños de cuadrados latinos

Cuando varios factores tienen que ser estudiados al mismo tiempo, un experimento en el cual existe por lo menos una observación por cada combinación posible de niveles se conoce como **diseño completo.** Si los factores son *A*, *B* y *C* con *I*, *J* y *K* niveles, respectivamente, un diseño completo requiere por lo menos *IJK* observaciones. Con frecuencia un experimento de este tamaño es impracticable debido a las restricciones de costo, tiempo o espacio o literalmente imposible. Por ejemplo, si la variable de respuesta es ventas de un cierto producto y los factores son configuraciones de exhibición diferentes, diferentes tiendas y diferentes lapsos de tiempo, entonces sólo una configuración de exhibición puede realísticamente ser usada en una tienda dada en un lapso de tiempo dado.

Un experimento con tres factores en el cual se realizan menos de *IJK* observaciones se llama **diseño incompleto.** Existen algunos diseños incompletos en los cuales el patrón de combinaciones de factores es tal que el análisis es directo. Un diseño de tres factores como ése se llama diseño de **cuadrado latino.** Es apropiado cuando $I = J = K$ (p. ej., cuatro configuraciones de exhibición, cuatro tiendas y cuatro lapsos de tiempo) y todos los efectos de interacción de dos y tres factores se suponen ausentes. Si los niveles del factor *A* están identificados con los renglones de una tabla bidireccional y los niveles de *B* con las columnas de la tabla, entonces la característica definitoria de un diseño de cuadrado latino es que *cada nivel del factor C aparece exactamente una vez en cada renglón y exactamente una vez en cada columna*. La figura 11.9 ilustra ejemplos de cuadrados latinos de 3×3, 4×4 y 5×5. Existen 12 cuadrados latinos diferentes de 3×3 y el número de cuadrados latinos diferentes se incrementa con rapidez con el número de niveles (p. ej., cada permutación de los renglones de un cuadrado latino dado produce un cuadrado latino y asimismo en el caso de permutaciones de columnas). Se recomienda que el cuadrado utilizado en realidad en un experimento particular se elija al azar del conjunto de todos los cuadrados posibles de la dimensión deseada, para más detalles, consulte una de las referencias del capítulo.

	B		
C ↘	1	2	3
1	1	2	3
A 2	2	3	1
3	3	1	2

	B			
C ↘	1	2	3	4
1	3	4	2	1
A 2	4	2	1	3
3	2	1	3	4
4	1	3	4	2

	B				
C ↘	1	2	3	4	5
1	4	3	5	2	1
2	3	1	4	5	2
A 3	1	5	2	3	4
4	5	2	1	4	3
5	2	4	3	1	5

Figura 11.9 Ejemplos de cuadrados latinos

Se utilizará la letra *N* para denotar el valor común de *I*, *J* y *K*. Entonces un diseño completo con una observación por cada combinación requeriría N^3 observaciones, en tanto que un cuadrado latino requiere sólo N^2 observaciones. Una vez que se selecciona un cuadrado particular, el valor de *k* (el nivel del factor *C*) queda determinado por completo por los valores de *i* y *j*. Para recalcar esto, se utiliza $x_{ij(k)}$ para denotar el valor observado cuando los tres factores están a los niveles *i*, *j* y *k*, respectivamente, con *k* tomando sólo un valor por cada par *i*, *j*.

La ecuación modelo para un diseño de cuadrado latino es

$$X_{ij(k)} = \mu + \alpha_i + \beta_j + \delta_k + \epsilon_{ij(k)} \quad i, j, k = 1, \ldots, N$$

donde $\sum\alpha_i = \sum\beta_j = \sum\delta_k = 0$ y las $\epsilon_{ij(k)}$ son independientes y normalmente distribuidas con media 0 y varianza σ^2.

Se emplea la siguiente notación para los totales y promedios:

$$X_{i\cdot\cdot} = \sum_j X_{ij(k)} \qquad X_{\cdot j\cdot} = \sum_i X_{ij(k)} \qquad X_{\cdot\cdot k} = \sum_{i,j} X_{ij(k)} \qquad X_{\cdots} = \sum_i \sum_j X_{ij(k)}$$

$$\bar{X}_{i\cdot\cdot} = \frac{X_{i\cdot\cdot}}{N} \qquad \bar{X}_{\cdot j\cdot} = \frac{X_{\cdot j\cdot}}{N} \qquad \bar{X}_{\cdot\cdot k} = \frac{X_{\cdot\cdot k}}{N} \qquad \bar{X}_{\cdots} = \frac{X_{\cdots}}{N^2}$$

Obsérvese que aunque $X_{i\cdot\cdot}$ previamente sugería una suma doble, ahora corresponde a una sola suma para todas las j (y los valores asociados de k).

DEFINICIÓN

Las **sumas de cuadrados** para un experimento de cuadrado latino son

$$\text{SST} = \sum_i \sum_j (X_{ij(k)} - \bar{X}_{\cdots})^2 \qquad \text{grados de libertad} = N^2 - 1$$

$$\text{SSA} = \sum_i \sum_j (\bar{X}_{i\cdot\cdot} - \bar{X}_{\cdots})^2 \qquad \text{grados de libertad} = N - 1$$

$$\text{SSB} = \sum_i \sum_j (X_{\cdot j\cdot} - \bar{X}_{\cdots})^2 \qquad \text{grados de libertad} = N - 1$$

$$\text{SSC} = \sum_i \sum_j (\bar{X}_{\cdot\cdot k} - \bar{X}_{\cdots})^2 \qquad \text{grados de libertad} = N - 1$$

$$\text{SSE} = \sum_i \sum_j [X_{ij(k)} - (\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\delta}_k)]^2 \qquad \text{grados de libertad} = N - 1$$

$$= \sum_i \sum_j (X_{ij(k)} - \bar{X}_{i\cdot\cdot} - \bar{X}_{\cdot j\cdot} - \bar{X}_{\cdot\cdot k} + 2\bar{X}_{\cdots})^2 \qquad \text{grados de libertad} = (N-1)(N-2)$$

$$\text{SST} = \text{SSA} + \text{SSB} + \text{SSC} + \text{SSE}$$

Cada media cuadrática es, por supuesto, la razón SS/gl. Para probar H_{0C}: $\delta_1 = \delta_2 = \cdots = \delta_N = 0$, el valor estadístico de prueba es $f_C = \text{MSC/MSE}$, con H_{0C} rechazada si $f_C \geq F_{\alpha,N-1,(N-1)(N-2)}$. Las otras dos hipótesis nulas para efectos principales también son rechazadas si la razón F correspondiente es al menos $F_{\alpha,N-1,(N-1)(N-2)}$.

Si cualquiera de las hipótesis nulas es rechazada, las diferencias significativas pueden ser identificadas por medio del procedimiento de Tukey. Después de calcular $w = Q_{\alpha,N,(N-1)(N-2)} \cdot \sqrt{\text{MSE}/N}$, los pares de medias muestrales (las $\bar{x}_{i\cdot\cdot}$, $\bar{x}_{\cdot j\cdot}$ o $\bar{x}_{\cdot\cdot k}$ que difieren por más de w corresponden a diferencias significativas entre los efectos del factor asociado (las α_i, β_j o δ_k).

La hipótesis H_{0C} es con frecuencia la de interés central. Se utiliza un diseño de cuadrado latino para controlar la variación externa de los factores A y B, como se hizo mediante un diseño de bloques aleatorizados en el caso de un factor externo único. Así pues en el ejemplo de ventas de productos previamente mencionado, la variación debida tanto a las tiendas como a los lapsos de tiempo es controlada por un diseño de cuadrado latino, lo que permite que un investigador pruebe en cuanto a la presencia de efectos producidos por las diferentes configuraciones de exhibición de productos.

Ejemplo 11.11 En un experimento para investigar el efecto de la humedad relativa en la resistencia a la abrasión de piel recortada de un patrón rectangular (“The Abrasion of Leather”, *J. Inter. Soc. Leather Trades’ Chemists*, 1946: 287), se utilizó un cuadrado latino de 6 × 6 para controlar la posible variabilidad a causa de la posición en los renglones y columnas del patrón. Los seis niveles de humedad relativa estudiados fueron 1 = 25%, 2 = 37%, 3 = 50%, 4 = 62%, 5 = 75% y 6 = 87% con los siguientes resultados:

		B (columnas)						
		1	**2**	**3**	**4**	**5**	**6**	$x_{i..}$
A (renglones)	**1**	$^{3}7.38$	$^{4}5.39$	$^{6}5.03$	$^{2}5.50$	$^{5}5.01$	$^{1}6.79$	35.10
	2	$^{2}7.15$	$^{1}8.16$	$^{5}4.96$	$^{4}5.78$	$^{3}6.24$	$^{6}5.06$	37.35
	3	$^{4}6.75$	$^{6}5.64$	$^{3}6.34$	$^{5}5.31$	$^{1}7.81$	$^{2}8.05$	39.90
	4	$^{1}8.05$	$^{3}6.45$	$^{2}6.31$	$^{6}5.46$	$^{4}6.05$	$^{5}5.51$	37.83
	5	$^{6}5.65$	$^{5}5.44$	$^{1}7.27$	$^{3}6.54$	$^{2}7.03$	$^{4}5.96$	37.89
	6	$^{5}6.00$	$^{2}6.55$	$^{4}5.93$	$^{1}8.02$	$^{6}5.80$	$^{3}6.61$	38.91
	$x_{.j.}$	40.98	37.63	35.84	36.61	37.94	37.98	

Además, $x_{..1} = 46.10$, $x_{..2} = 40.59$, $x_{..3} = 39.56$, $x_{..4} = 35.86$, $x_{..5} = 32.23$, $x_{..6} = 32.64$, $x_{...} = 226.98$. En la tabla 11.9 aparecen más cálculos.

Tabla 11.9 Tabla ANOVA para el ejemplo 11.11

Causa de la variación	Grados de libertad	Suma de cuadrados	Medias cuadráticas	*f*
A (renglones)	5	2.19	.438	2.50
B (columnas)	5	2.57	.514	2.94
C (tratamientos)	5	23.53	4.706	26.89
Error	20	3.49	.175	
Total	35	31.78		

Puesto que $F_{.05,5,20} = 2.71$ y $26.89 \geq 2.71$, H_{0C}, es rechazada a favor de la hipótesis de que la humedad relativa sí afecta en promedio la resistencia a la abrasión.

Para aplicar el procedimiento de Tukey, $w = Q_{.05,6,20} \cdot \sqrt{\text{MSE}/6} = 4.45\sqrt{.175/6} = .76$. Después de ordenar las $\bar{x}_{..k}$ y subrayarlas se obtiene

75%	87%	62%	50%	37%	25%
5.37	5.44	5.98	6.59	6.77	7.68

En particular, la humedad relativa más baja aparentemente produce una resistencia a la abrasión promedio verdadera significativamente más alta que cualquier otra humedad relativa estudiada. ■

EJERCICIOS Sección 11.3 (27–37)

27. Se estudió el rendimiento de una máquina de extrusión continua que recubre tubos de acero con plástico como una función del perfil de temperatura del termostato (*A*, a tres niveles), tipo de plástico (*B*, a tres niveles) y la velocidad de tornillo rotatorio que hace que el plástico pase a través del troquel formador de tubos (*C*, a tres niveles). Hubo dos réplicas ($L = 2$) con cada combinación de niveles de los factores lo que produjo un total de 54 observaciones del rendimiento. Las sumas de cuadrados fueron SSA = 14,144.44, SSB = 5511.27, SSC = 244,696.39, SSAB = 1069.62, SSAC = 62.67, SSBC = 331.67, SSE = 3127.50 y SST = 270,024.33.

a. Construya la tabla ANOVA.

b. Use pruebas *F* apropiadas para demostrar que ninguna de las razones *F* para interacción de dos o tres factores es significativa al nivel .05.

c. ¿Qué efectos principales parecen significativos?

d. Con $x_{..1.} = 8242$, $x_{..2.} = 9732$ y $x_{..3.} = 11{,}210$, use el procedimiento de Tukey para identificar diferencias significativas entre los niveles del factor *C*.

28. Para ver si la fuerza de empuje al taladrar es afectada por la velocidad de taladrado (*A*), coeficiente de alimentación (*B*) o material utilizado (*C*), se realizó un experimento utilizando cuatro velocidades, tres coeficientes y dos materiales, con dos muestras ($L = 2$) taladradas con cada combinación de niveles de los tres factores). Las sumas de cuadrados se calcularon como sigue: SSA = 19,149.73, SSB = 2,589,047.62, SSC = 157,437.52, SSAB = 53,238.21, SSAC = 9033.73, SSBC = 91,880.04, SSE = 56,819.50, y SST = 2,983,164.81. Construya la tabla ANOVA e identifique interacciones significativas con

$\alpha = .01$. ¿Existe algún factor que parezca no tener efecto en la fuerza de empuje? (En otras palabras, ¿Parece ser algún factor no significativo en todo efecto en el que aparece?

29. El artículo "An Analysis of Variance Applied to Screw Machines" (*Industrial Quality Control*, 1956: 8–9) describe un experimento para investigar cómo la longitud de barras de acero se vio afectada por la hora del día (A), el tratamiento térmico aplicado (B) y la máquina de roscar utilizada (C). Las tres horas fueron 8:00 a.m., 11:00 a.m. y 3:00 p.m. y hubo dos tratamientos y cuatro máquinas (un experimento factorial de $3 \times 2 \times 4$) y se obtuvieron los datos adjuntos [codificados como 1000(longitud − 4.380), lo cual no afecta el análisis].

B_1

	C_1	C_2	C_3	C_4
A_1	6, 9, 1, 3	7, 9, 5, 5	1, 2, 0, 4	6, 6, 7, 3
A_2	6, 3, 1, −1	8, 7, 4, 8	3, 2, 1, 0	7, 9, 11, 6
A_3	5, 4, 9, 6	10, 11, 6, 4	−1, 2, 6, 1	10, 5, 4, 8

B_2

	C_1	C_2	C_3	C_4
A_1	4, 6, 0, 1	6, 5, 3, 4	−1, 0, 0, 1	4, 5, 5, 4
A_2	3, 1, 1, −2	6, 4, 1, 3	2, 0, −1, 1	9, 4, 6, 3
A_3	6, 0, 3, 7	8, 7, 10, 0	0, −2, 4, −4	4, 3, 7, 0

Las sumas de los cuadrados incluyen SSAB = 1.646, SSAC = 71.021, SSBC = 1.542, SSE = 447.500 y SST = 1037.833.

a. Construya la tabla ANOVA con estos datos.
b. Pruebe para ver si algunos efectos de interacción son significativos al nivel .05.
c. Pruebe para ver si alguno de los efectos principales son significativos al nivel .05 (es decir, H_{0A} contra H_{aA}, etc.).
d. Use el procedimiento de Tukey para investigar diferencias significativas entre las cuatro máquinas.

30. Se calcularon las siguientes cantidades con un experimento que implicó cuatro niveles de nitrógeno (A), dos tiempos de plantación (B) y dos niveles de potasio (C) ("Use and Misuse of Multiple Comparison Procedures", *Agronomy J.*, 1977: 205–208). Se realizó sólo una observación (contenido de N, en porcentaje, de granos de maíz) por cada una de las 16 combinaciones de niveles.

SSA = .22625 SSB = .000025 SSC = .0036
SSAB = .004325 SSAC = .00065
SSBC = .000625 SST = .2384.

a. Construya la tabla ANOVA.
b. Suponga que no existen efectos de interacción en tres direcciones, de modo que MSABC es una estimación válida de σ^2 y pruebe al nivel .05 en cuanto a interacción y efectos principales.
c. Los promedios de nitrógeno son $\bar{x}_{1..} = 1.1200$, $\bar{x}_{2..} = 1.3025$, $\bar{x}_{3..} = 1.3875$ y $\bar{x}_{4..} = 1.4300$. Use el método de Tukey para examinar las diferencias de porcentaje N entre los niveles de nitrógeno ($Q_{.05,4,3} = 6.82$).

31. El artículo "Kolbe–Schmitt Carbonation of 2-Naphthol" (*Industrial and Engr. Chemistry*: *Process and Design Development*, 1969: 165–173) presentó los datos adjuntos sobre porcentaje de rendimiento de ácido BON en función del tiempo de reacción (1, 2 y 3 horas), temperatura (30, 70 y 100°C) y presión (30, 70 y 100 lb/pulg^2). Suponiendo que no existe interacción de tres factores, de modo que SSE = SSABC proporcione una estimación de σ^2. Minitab dio la tabla ANOVA adjunta. Realice todas las pruebas apropiadas.

B_1

	C_1	C_2	C_3
A_1	68.5	73.0	68.7
A_2	74.5	75.0	74.6
A_3	70.5	72.5	74.7

B_2

	C_1	C_2	C_3
A_1	72.8	80.1	72.0
A_2	72.0	81.5	76.0
A_3	69.5	84.5	76.0

B_3

	C_1	C_2	C_3
A_1	72.5	72.5	73.1
A_2	75.5	70.0	76.0
A_3	65.0	66.5	70.5

Análisis de varianza para rendimiento

Fuente	DF	SS	MS	F	P
tiempo	2	42.112	21.056	8.76	0.010
temperatura	2	110.732	55.366	23.04	0.000
presión	2	68.136	34.068	14.18	0.002
tiempo* temperatura	4	67.761	16.940	7.05	0.010
tiempo* presión	4	35.184	8.796	3.66	0.056
temperatura* presión	4	136.437	34.109	14.20	0.001
Error	8	19.223	2.403		
Total	26	479.585			

32. Cuando los factores A y B son fijos pero el factor C es aleatorio y se utiliza el modelo restringido (véase la nota al pie de página 438; existe una complicación técnica con el modelo no restringido en este caso), y $E(\text{MSE}) = \sigma^2$

$$E(\text{MSA}) = \sigma^2 + JL\sigma^2_{AC} + \frac{JKL}{I-1}\sum\alpha_i^2$$

$$E(\text{MSB}) = \sigma^2 + IL\sigma^2_{BC} + \frac{IKL}{J-1}\sum\beta_j^2$$

$$E(\text{MSC}) = \sigma^2 + IJL\sigma^2_C$$

$$E(\text{MSAB}) = \sigma^2 + L\sigma^2_{ABC} + \frac{KL}{(I-1)(J-1)}\sum_i\sum_j(\gamma_{ij}^{AB})^2$$

$$E(\text{MSAC}) = \sigma^2 + JL\sigma^2_{AC}$$

$$E(\text{MSBC}) = \sigma^2 + IL\sigma^2_{BC}$$

$$E(\text{MSABC}) = \sigma^2 + \sigma^2_{ABC}$$

a. Basado en estas medias cuadráticas esperadas, ¿qué razones F utilizaría para probar H_0: $\sigma^2_{ABC} = 0$; H_0: $\sigma^2_C = 0$; H_0: $\gamma_{ij}^{AB} = 0$ para todas las i, j; y H_0: $\alpha_1 = \cdots = \alpha_I = 0$?

b. En un experimento para valuar los efectos de la edad, el tipo de suelo y el día de la producción en la resistencia a la compresión de mezclas de cemento/suelo, se utilizaron dos edades (A), cuatro tipos de suelo (B) y 3 días (C, supuesto aleatorio), con $L = 2$ observaciones realizadas por cada combinación de niveles de factor. Las sumas de cuadrados resultantes fueron SSA = 14,318.24, SSB = 9656.40, SSC = 2270.22, SSAB = 3408.93, SSAC = 1442.58, SSBC = 3096.21, SSABC = 2832.72, y SSE = 8655.60. Obtenga la tabla ANOVA y realice todas las pruebas al nivel .01.

33. Debido a la variabilidad potencial del envejecimiento causado por las diferentes piezas fundidas y segmentos en éstas, se utilizó un diseño de cuadrado latino con $N = 7$ para investigar el efecto del tratamiento térmico en el envejecimiento. Con A = piezas fundidas, B = segmentos, C = tratamientos térmicos, los estadísticos resumidos incluyen $x_{\cdots} = 3815.8$, $\sum x_{i\cdot\cdot}^2 = 297{,}216.90$, $\sum x_{\cdot j\cdot}^2 = 297{,}200.64$, $\sum x_{\cdot\cdot k}^2 = 297{,}155.01$, y $\sum\sum x_{ij(k)}^2 = 297{,}317.65$. Obtenga la tabla ANOVA y pruebe al nivel .05 la hipótesis de que el tratamiento térmico no afecta el envejecimiento.

34. El artículo "The Responsiveness of Food Sales to Shelf Space Requirements" (*J. Marketing Research*, 1964: 63–67) reporta el uso de un diseño de cuadrado latino para investigar el efecto del espacio de anaquel en las ventas de alimentos. El experimento se realizó a lo largo de un periodo de 6 semanas con seis tiendas diferentes y se obtuvieron los siguientes resultados sobre ventas de crema en polvo para café (con el índice de espacio de anaquel entre paréntesis):

		Semana		
		1	**2**	**3**
Tienda	**1**	27 (5)	14 (4)	18 (3)
	2	34 (6)	31 (5)	34 (4)
	3	39 (2)	67 (6)	31 (5)
	4	40 (3)	57 (1)	39 (2)
	5	15 (4)	15 (3)	11 (1)
	6	16 (1)	15 (2)	14 (6)

		Semana		
		4	**5**	**6**
Tienda	**1**	35 (1)	28 (6)	22 (2)
	2	46 (3)	37 (2)	23 (1)
	3	49 (4)	38 (1)	48 (3)
	4	70 (6)	37 (4)	50 (5)
	5	9 (2)	18 (5)	17 (6)
	6	12 (5)	19 (3)	22 (4)

Construya la tabla ANOVA y formule y pruebe al nivel .01 la hipótesis de que el espacio de anaquel no afecta las ventas contra la alternativa apropiada.

35. El artículo "Variation in Moisture and Ascorbic Acid Content from Leaf to Leaf and Plant to Plant in Turnip Greens" (*Southern Cooperative Services Bull.*, 1951: 13–17) use un diseño de cuadrado latino en el cual el factor A es la planta, el factor B es el tamaño de hoja (desde el más pequeño hasta el más grande), el factor C (entre paréntesis) es tiempo de pesada y la variable de respuesta es el contenido de humedad.

		Tamaño de hoja (B)		
		1	**2**	**3**
Planta (A)	**1**	6.67 (5)	7.15 (4)	8.29 (1)
	2	5.40 (2)	4.77 (5)	5.40 (4)
	3	7.32 (3)	8.53 (2)	8.50 (5)
	4	4.92 (1)	5.00 (3)	7.29 (2)
	5	4.88 (4)	6.16 (1)	7.83 (3)

		Tamaño de hoja (B)	
		4	**5**
Planta (A)	**1**	8.95 (3)	9.62 (2)
	2	7.54 (1)	6.93 (3)
	3	9.99 (4)	9.68 (1)
	4	7.85 (5)	7.08 (4)
	5	5.83 (2)	8.51 (5)

Cuando los tres factores son aleatorios, las medias cuadráticas esperadas son $E(\text{MSA}) = \sigma^2 + N\sigma^2_A$, $E(\text{MSB}) = \sigma^2 + N\sigma^2_B$, $E(\text{MSC}) = \sigma^2 + N\sigma^2_C$ y $E(\text{MSE}) = \sigma^2$. Esto implica que las razones F para probar H_{0A}: $\sigma^2_A = 0$, H_{0B}: $\sigma^2_B = 0$ y H_{0C}: $\sigma^2_C = 0$ son idénticas a aquellas para efectos fijos. Obtenga la tabla ANOVA y pruebe al nivel .05 para ver si existe alguna variación en el contenido de humedad debido a los factores.

36. El artículo "An Assessment of the Effects of Treatment, Time and Heat on the Removal of Erasable Pen Marks from Cotton and Cotton/Poliester Blend Fabrics" (*J. of Testing and Eval.*, 1991: 394–397) reporta las siguientes sumas de cuadrados para la variable de respuesta *grado de eliminación de marcas*: SSA = 39.171, SSB = .665, SSC = 21.508, SSAB = 1.432, SSAC = 15.953, SSBC = 1.382, SSABC = 9.016, y SSE = 115.820. Se utilizaron cuatro tratamientos de lavado diferentes, tres tipos diferentes de pluma y seis telas diferentes en el experimento y se realizaron tres observaciones por cada combinación de pluma-tela. Analice la

varianza con $\alpha = .01$ por cada prueba y exprese sus conclusiones (suponga efectos fijos para los tres factores).

37. Se realizó un experimento ANOVA con cuatro factores para investigar los efectos de la tela (A), el tipo de exposición (B), el nivel de exposición (C) y la dirección de la tela (D) en el grado de cambio de color en telas expuestas medido por medio de un espectrocolorímetro. Se realizaron dos observaciones por cada una de las tres telas, dos tipos, tres niveles y dos direcciones con los siguientes resultados: MSA = 2207.329, MSB = 47.255, MSC = 491.783, MSD = .044, MSAB = 15.303, MSAC = 275.446, MSAD = .470, MSBC = 2.141, MSBD = .273, MSCD = .247, MSABC = 3.714, MSABD = 4.072, MSABD = 4.072, MSACD = .767, MSBCD = .280, MSE = .977, y MST = 93.621. ("Accelerated Weathering of Marine Fabrics", *J. Testing and Eval.*, 1992: 139–143). Suponiendo efectos fijos con todos los factores, analice la varianza con $\alpha = .01$ con todas las pruebas y resuma sus conclusiones.

11.4 Experimentos 2^p factoriales

Si un experimentador desea estudiar al mismo tiempo el efecto de p factores diferentes en una variable de respuesta y los factores tienen $I_1, I_2, \ldots, I_p$ niveles, respectivamente, entonces un experimento completo requiere por lo menos $I_1 \cdot I_2, \cdot \cdots \cdot I_p$ observaciones. En tales situaciones, el experimentador a menudo puede realizar un "experimento de filtración" con cada factor a sólo dos niveles para obtener información preliminar sobre los efectos del factor. Un experimento en el cual existen p factores, cada uno a dos niveles, se conoce como **experimento 2^p factorial.**

Experimentos 2^3

Como en la sección 11.3, X_{ijkl} y x_{ijkl} se refieren a la observación de la l-ésima réplica con los factores A, B y C a los niveles i, j y k, respectivamente. El modelo en esta situación es

$$X_{ijkl} = \mu + \alpha_i + \beta_j + \delta_k + \gamma_{ij}^{AB} + \gamma_{ik}^{AC} + \gamma_{jk}^{BC} + \gamma_{ijk} + \epsilon_{ijkl} \qquad (11.14)$$

para $i = 1, 2$; $j = 1, 2$; $k = 1, 2$; $l = 1, \ldots, n$. Las ϵ_{ijkl} se suponen independientes, normalmente distribuidas, con media 0 y varianza σ^2. Como existen sólo dos niveles de cada factor, las condiciones laterales en relación con los parámetros de (11.14) que especifican de manera única el modelo simplemente se formulan como $\alpha_1 + \alpha_2 = 0, \ldots,$ $\gamma_{11}^{AB} + \gamma_{21}^{AB} = 0, \gamma_{12}^{AB} + \gamma_{22}^{AB} = 0,$ $\gamma_{11}^{AB} + \gamma_{12}^{AB} = 0, \gamma_{21}^{AB} + \gamma_{22}^{AB} = 0$ y similares. Estas condiciones implican que existe sólo un parámetro funcionalmente independiente de cada tipo (por cada efecto principal e interacción). Por ejemplo, $\alpha_2 = -\alpha_1$, mientras que $\gamma_{21}^{AB} = -\gamma_{11}^{AB}$, $\gamma_{12}^{AB} = -\gamma_{11}^{AB}$, y $\gamma_{22}^{AB} = \gamma_{11}^{AB}$. Debido a esto, cada suma de cuadrados en el análisis tendrá 1 grado de libertad.

Los parámetros del modelo pueden ser estimados sacando promedios para todos los subíndices de las X_{ijkl} y luego formando combinaciones lineales apropiadas de los promedios. Por ejemplo,

$$\begin{aligned}\hat{\alpha}_1 &= \bar{X}_{1\cdots} - \bar{X}_{\cdots} \\ &= \frac{(X_{111\cdot} + X_{121\cdot} + X_{112\cdot} + X_{122\cdot} - X_{211\cdot} - X_{212\cdot} - X_{221\cdot} - X_{222\cdot})}{8n}\end{aligned}$$

y

$$\begin{aligned}\hat{\gamma}_{11}^{AB} &= \bar{X}_{11\cdot\cdot} - \bar{X}_{1\cdots} - \bar{X}_{\cdot 1\cdot\cdot} + \bar{X}_{\cdots} \\ &= \frac{(X_{111\cdot} - X_{121\cdot} - X_{211\cdot} + X_{221\cdot} + X_{112\cdot} - X_{122\cdot} - X_{212\cdot} + X_{222\cdot})}{8n}\end{aligned}$$

Cada estimador es, con excepción del factor $1/(8n)$, una función lineal de los totales de celda ($X_{ijk\cdot}$) donde cada coeficiente es $+1$ o -1, con un número igual de cada uno; tales

funciones se llaman **contrastes** en las X_{ijk}. Además, los estimadores satisfacen las mismas condiciones laterales satisfechas por los parámetros mismos. Por ejemplo,

$$\hat{\alpha}_1 + \hat{\alpha}_2 = \overline{X}_{1\ldots} - \overline{X}_{\ldots} + \overline{X}_{2\ldots} - \overline{X}_{\ldots} = \overline{X}_{1\ldots} + \overline{X}_{2\ldots} - 2\overline{X}_{\ldots}$$
$$= \frac{1}{4n}X_{1\ldots} + \frac{1}{4n}X_{2\ldots} - \frac{2}{8n}X_{\ldots} = \frac{1}{4n}X_{\ldots} - \frac{1}{4n}X_{\ldots} = 0$$

Ejemplo 11.12 En un experimento para investigar las propiedades de resistencia a la compresión de mezclas de cemento–tierra, se utilizaron dos periodos de añejamiento en combinación con dos temperaturas diferentes y dos tierras distintas. Se hicieron dos réplicas por cada combinación de niveles de los tres factores y se obtuvieron los siguientes resultados:

		Suelo	
Añejamiento	**Temperatura**	**1**	**2**
1	**1**	471, 413	385, 434
	2	485, 552	530, 593
2	**1**	712, 637	770, 705
	2	712, 789	741, 806

Los totales de celda calculados son $x_{111.} = 884$, $x_{211.} = 1349$, $x_{121.} = 1037$, $x_{221.} = 1501$, $x_{112.} = 819$, $x_{212.} = 1475$, $x_{122.} = 1123$ y $x_{222.} = 1547$, por lo tanto $x_{\ldots.} = 9735$. Entonces

$$\hat{\alpha}_1 = (884 - 1349 + 1037 - 1501 + 819 - 1475 + 1123 - 1547)/16$$
$$= -125.5625 = -\hat{\alpha}_2$$
$$\hat{\gamma}_{11}^{AB} = (884 - 1349 - 1037 + 1501 + 819 - 1475 - 1123 + 1547)/16$$
$$= -14.5625 = -\hat{\gamma}_{12}^{AB} = -\hat{\gamma}_{21}^{AB} = \hat{\gamma}_{22}^{AB}$$

Las estimaciones de los demás parámetros se calculan de la misma manera. ■

Análisis de un experimento 2^3 Las sumas de cuadrados para los varios efectos son fáciles de obtener a partir de los parámetros estimados. Por ejemplo,

$$\text{SSA} = \sum_i \sum_j \sum_k \sum_l \hat{\alpha}_i^2 = 4n \sum_{i=1}^{2} \hat{\alpha}_i^2 = 4n[\hat{\alpha}_1^2 + (-\hat{\alpha}_1)^2] = 8n\hat{\alpha}_1^2$$

y

$$\text{SSAB} = \sum_i \sum_j \sum_k \sum_l (\hat{\gamma}_{ij}^{AB})^2$$
$$= 2n \sum_{i=1}^{2} \sum_{j=1}^{2} (\hat{\gamma}_{ij}^{AB})^2 = 2n[(\hat{\gamma}_{11}^{AB})^2 + (-\hat{\gamma}_{11}^{AB})^2 + (-\hat{\gamma}_{11}^{AB})^2 + (\hat{\gamma}_{11}^{AB})^2]$$
$$= 8n(\hat{\gamma}_{11}^{AB})^2$$

Como cada estimación es un contraste en los totales de celda multiplicado por $1/(8n)$, cada suma de cuadrados tiene la forma $(\text{contraste})^2/(8n)$. Por lo tanto para calcular las diversas sumas de cuadrados, se tienen que conocer los coeficientes $(+1 \text{ o } -1)$ de los contrastes apropiados. Los signos $(+ \text{ o } -)$ de cada $x_{ijk.}$ en cada contraste de efecto son más convenientemente mostrados en una tabla. Se utilizará la notación (1) para la condición experimental $i = 1, j = 1, k = 1$, a para $i = 2, j = 1, k = 1$, ab para $i = 2, j = 2, k = 1$, y así sucesivamente. Si el nivel 1 se considera como "bajo" y el nivel 2 como "alto", cualquier letra que aparezca denota un nivel alto del factor asociado. En la tabla 11.10, cada columna da los signos para un contraste de efecto particular en las x_{ijk} asociadas con las diferentes condiciones experimentales.

Tabla 11.10 Signos para calcular contrastes de efecto

Condición experimental	Total de celda	Efecto factorial A	B	C	AB	AC	BC	ABC
(1)	$x_{111.}$	−	−	−	+	+	+	−
a	$x_{211.}$	+	−	−	−	−	+	+
b	$x_{121.}$	−	+	−	−	+	−	+
ab	$x_{221.}$	+	+	−	+	−	−	−
c	$x_{112.}$	−	−	+	+	−	−	+
ac	$x_{212.}$	+	−	+	−	+	−	−
bc	$x_{122.}$	−	+	+	−	−	+	−
abc	$x_{222.}$	+	+	+	+	+	+	+

En cada una de las tres primeras columnas, el signo es + si el factor correspondiente está al nivel alto y − si está al nivel bajo. Cada signo que aparece en la columna AB es entonces el "producto" de los signos presentes en las columnas A y B *con* $(+)(+) = (-)(-) = +$ y $(+)(-) = (-)(+) = -$ y del mismo modo para las columnas AC y BC. Por último, los signos que aparecen en la columna ABC son los productos de AB con C (o B con AC o A con BC). Así pues, por ejemplo,

$$AC \text{ contraste} = + x_{111.} - x_{211.} + x_{121.} - x_{221.} - x_{112.} + x_{212.} - x_{122.} + x_{222.}$$

Una vez que se calculan los siete contrastes de efecto,

$$\text{SS(efecto)} = \frac{(\text{contraste de efecto})^2}{8n}$$

El software para hacer los cálculos necesarios para analizar datos de experimentos factoriales está ampliamente disponible (por ejemplo, Minitab). De manera alternativa, existe un método eficiente de cálculo manual creado por Yates. Se escriben en una columna los ocho totales de celda en el **orden estándar** como aparece en la tabla de signos y se establecen tres columnas adicionales. En cada una de estas tres columnas, las primeras cuatro entradas son las sumas de las entradas 1 y 2, 3 y 4, 5 y 6, 7 y 8 de las columnas previas. Las últimas cuatro entradas son las diferencias entre las entradas 2 y 1, 4 y 3, 6 y 5 y 8 y 7 de la columna previa. La última columna contiene entonces x.... y los siete contrastes de efecto en orden estándar. Si se eleva al cuadrado cada contraste y se divide entre $8n$ se obtienen entonces las siete sumas de cuadrados.

Ejemplo 11.13 (Continuación del ejemplo 11.12)

Como $n = 2$, $8n = 16$, el método de Yates se ilustra en la tabla 11.11.

Tabla 11.11 Método de Yates de cálculo

Condición de tratamiento	$x_{ijk.}$	1	2	Contraste de efecto	$SS = (\text{contraste})^2/16$
$(1) = x_{111.}$	884	2233	4771	9735	
$a = x_{211.}$	1349	2538	4964	2009	252,255.06
$b = x_{121.}$	1037	2294	929	681	28,985.06
$ab = x_{221.}$	1501	2670	1080	−233	3,393.06
$c = x_{112.}$	819	465	305	193	2,328.06
$ac = x_{212.}$	1475	464	376	151	1,425.06
$bc = x_{122.}$	1123	656	−1	71	315.06
$abc = x_{222.}$	1547	424	−232	−231	3,335.06
					292,036.42

Con los datos originales, $\sum_i\sum_j\sum_k\sum_l x_{ijkl}^2 = 6{,}232{,}289$, y

$$\frac{x_{....}^2}{16} = 5{,}923{,}139.06$$

por lo tanto

$$\begin{aligned} \text{SST} &= 6{,}232{,}289 - 5{,}923{,}139.06 = 309{,}149.94 \\ \text{SSE} &= \text{SST} - [\text{SSA} + \cdots + \text{SSABC}] = 309{,}149.94 - 292{,}036.42 \\ &= 17{,}113.52 \end{aligned}$$

Los cálculos ANOVA se resumen en la tabla 11.12.

Tabla 11.12 Tabla ANOVA para el ejemplo 11.13

Causa de la variación	Grados de libertad	Suma de cuadrados	Medias cuadráticas	*f*
A	1	252,255.06	252,255.06	117.92
B	1	28,985.06	28,985.06	13.55
C	1	2,328.06	2,328.06	1.09
AB	1	3,393.06	3,393.06	1.59
AC	1	1,425.06	1,425.06	.67
BC	1	315.06	315.06	.15
ABC	1	3,335.06	3,335.06	1.56
Error	8	17,113.52	2,139.19	
Total	15	309,149.94		

La figura 11.10 muestra los resultados generados por SAS para este ejemplo. Sólo los valores *P* correspondientes a la edad (*A*) y temperatura (*B*) son menores que .01, así que sólo estos efectos son considerados significativos.

```
Análisis del proceso de varianza
Variable dependiente: RESISTENCIA

                              Suma de          Media
Fuente              GL      cuadrados    cuadrática    Valor F    Pr > F
Modelo               7    292036.4375   41719.4911      19.50    0.0002
Error                8     17113.5000    2139.1875
Total corregido     15    309149.9375

                                                          Media de
                  Raíz-R          C.V.       Raíz MSE     USO DE ENERGIA

                0.944643      7.601660       46.25135          608.437500

                                                 Media
Fuente                GL      Anova SS    cuadrática    Valor F   Pr > F

AÑEJAMIENTO            1   252255.0625   252255.0625     117.92   0.0001
TEMPERATURA            1    28985.0625    28985.0625      13.55   0.0062
AÑE*TEMPERATURA        1     3393.0625     3393.0625       1.59   0.2434
SUELO                  1     2328.0625     2328.0625       1.09   0.3273
AÑE*SUELO              1     1425.0625     1425.0625       0.67   0.4380
TEMPERATURA*SUELO      1      315.0625      315.0625       0.15   0.7111
AÑE*TEMPERATURA*SUELO  1     3335.0625     3335.0625       1.56   0.2471
```

Figura 11.10 Resultados obtenidos con SAS con los datos de resistencia del ejemplo 11.13 ■

Experimentos 2^p con $p > 3$

El análisis de los datos de un experimento 2^p con $p > 3$ es paralelo al del caso de tres factores. Por ejemplo, si existen cuatro factores *A*, *B*, *C* y *D*, existen 16 condiciones experimentales diferentes. Las primeras 8 en orden estándar son exactamente las que ya aparecen en lista para un experimento con tres factores. Las segundas 8 se obtienen colocando la

letra *d* al lado de cada condición en el primer grupo. El método de Yates se inicia entonces calculando totales a través de las réplicas, poniendo en lista estos totales en orden estándar y procediendo como antes; con p factores, la p-ésima columna a la derecha de los totales de tratamiento dará los contrastes de efecto.

Con $p > 3$, con frecuencia no habrá réplicas del experimento (así que sólo una réplica completa está disponible). Una posible forma de probar hipótesis es asumir que ciertos efectos de alto grado están ausentes y luego agregar las sumas correspondientes de cuadrados para obtener un SSE. Tal suposición, sin embargo, puede ser engañosa si no se tiene un conocimiento previo (véase el libro de Montgomery que aparece en la bibliografía del capítulo). Un método alternativo implica trabajar directamente con los contrastes de efecto. Cada contraste tiene una distribución normal con la misma varianza. Cuando un efecto particular está ausente, el valor esperado del contraste correspondiente es 0, pero esto no es así cuando el efecto está presente. El método de análisis sugerido es construir una gráfica de probabilidad normal de los contrastes de efecto (o, de forma equivalente, las estimaciones de los parámetros de efecto, puesto que estimación = contraste/2^p cuando $n = 1$). Los puntos correspondientes a efectos ausentes tenderán a acercarse a una línea recta, mientras que los puntos asociados con efectos sustanciales en general se alejarán de esta línea.

Ejemplo 11.14 Los datos adjuntos se tomaron del artículo "Quick and Easy Analysis of Unreplicated Factorials" (*Technometrics*, 1989: 469–473). Los cuatro factores son A = resistencia al ácido, B = tiempo, C = cantidad de ácido y D = temperatura y la variable de respuesta es el rendimiento de isatina. Las observaciones, en orden estándar, son .08, .04, .53, .43, .31, .09, .12, .36, .79, .68, .73, .08, .77, .38, .49 y .23. La tabla 11.13 muestra las estimaciones de efecto como aparecen en el artículo (las cuales utilizaron constraste/8 en lugar de contraste/16).

Tabla 11.13 Estimaciones de efecto para el ejemplo 11.14

Efecto	*A*	*B*	*AB*	*C*	*AC*	*BC*	*ABC*	*D*
estimación	−.191	−.021	−.001	−.076	.034	−.066	.149	.274
Efecto	*AD*	*BD*	*ABD*	*CD*	*ACD*	*BCD*	*ABCD*	
estimación	−.161	−.251	−.101	−.026	−.066	.124	.019	

La figura 11.11 es una gráfica de probabilidad normal de las estimaciones de efecto. Todos los puntos en la gráfica quedan cerca de la misma línea recta, lo que sugiere la ausencia completa de cualquier efecto (en breve se dará un ejemplo en el cual éste no es el caso).

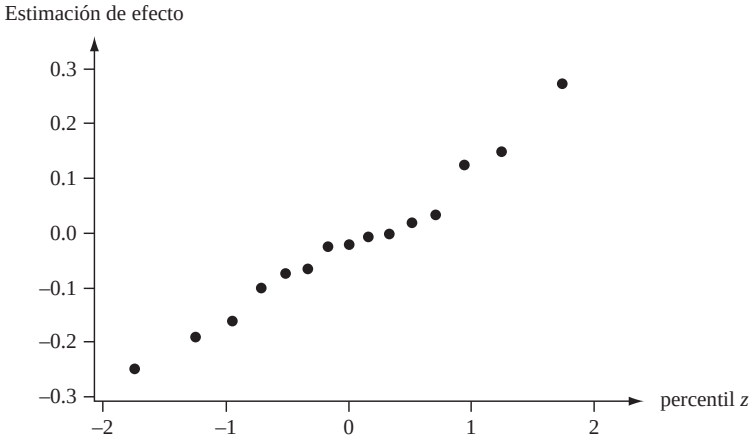


Figura 11.11 Curva de probabilidad normal de estimaciones de efecto del ejemplo 11.14 ■

Los juicios visuales de la desviación lineal en una gráfica de probabilidad normal son más bien subjetivos. El artículo citado en el ejemplo 11.14 describe una técnica más objetiva de identificar efectos significativos en un experimento no replicado.

Confusión

A menudo no es posible realizar todas las 2^p condiciones experimentales de un experimento 2^r factorial en un entorno experimental homogéneo. En tales situaciones, puede ser posible separar las condiciones experimentales en 2^r bloques homogéneos ($r < p$), de modo que existen 2^{p-r} condiciones experimentales en cada bloque. Los bloques pueden, por ejemplo, corresponder a laboratorios diferentes, lapsos de tiempo diferentes u operadores o cuadrillas de trabajo diferentes. En el caso más simple, $p = 3$ y $r = 1$, de modo que existen dos bloques con cada uno compuesto de cuatro de las ocho condiciones experimentales.

Como siempre, la formación de bloques es efectiva al reducir la variación asociada con fuentes externas. Sin embargo, cuando las 2^p condiciones experimentales se colocan en 2^r bloques, el precio pagado por esta formación de bloques es que $2^r - 1$ de los efectos de factor no pueden ser estimados. Esto es porque los $2^r - 1$ efectos de factor (efectos principales y/o interacciones) se mezclan o **confunden** con los efectos de bloque. La asignación de condiciones experimentales a bloques normalmente se hace de modo que sólo las interacciones de más alto nivel sean confundidas, mientras que los efectos principales y las interacciones de orden más bajo permanecen estimables y las hipótesis pueden ser probadas.

Para ver cómo se logra la asignación de bloques, considérese primero un experimento 2^3 con dos bloques ($r = 1$) y cuatro tratamientos por bloque. Supóngase que se selecciona *ABC* como el efecto que ha de ser confundido con bloques. Entonces cualquier condición experimental que tenga un número impar de letras en común con *ABC*, tal como *b* (una letra) o *abc* (tres letras) se coloca en un bloque, mientras que cualquier condición que tenga un número par de letras en común con *ABC* (donde 0 es par) va en el otro bloque. La figura 11.12 muestra esta asignación de tratamientos a los dos bloques.

Bloque 1	*Bloque 2*
(1), *ab*, *ac*, *bc*	*a*, *b*, *c*, *abc*

Figura 11.12 Confusión de *ABC* en un experimento 2^3.

Sin réplicas, los datos de semejante experimento normalmente se analizarían suponiendo que no hubo interacciones de dos factores (aditividad) y utilizando SSE = SSAB + SSAC + SSBC con 3 grados de libertad para probar en cuanto a la presencia de efectos principales. Alternativamente, una gráfica de probabilidad normal de contrastes de efecto o estimaciones de parámetros de efecto podría ser examinada. Con más frecuencia, no obstante, existen réplicas cuando sólo tres factores están siendo estudiados. Supóngase que existen *u* réplicas, que dan un total de $2^r \cdot u$ bloques en el experimento. Entonces después de restar de SST todas las sumas de cuadrados asociadas con efectos no confundidos con bloques (calculados con el método de Yates), el bloque de la suma de cuadrados se calcula con los $2^r \cdot u$ totales de bloque y luego se restan para obtener SSE (de modo que existen $2^r \cdot u - 1$ grados de libertad para los bloques).

Ejemplo 11.15 El artículo "Factorial Experiments in Pilot Plant Studies" (*Industrial and Eng. Chemistry*, 1951: 1300–1306) reporta los resultados de un experimento para valuar los efectos de temperatura de reactor (*A*), rendimiento de gas (*B*) y concentración de constituyente activo (*C*) en la concentración de la solución producto (medida en unidades arbitrarias) en una unidad de

recirculación. Se utilizaron dos bloques, con el efecto *ABC* confundido con bloques y hubo dos réplicas y los resultados aparecen en la figura 11.13. Los cuatro totales de bloque × réplica son 288, 212, 88 y 220 con un gran total de 808, por lo tanto

$$\text{SSB1} = \frac{(288)^2 + (212)^2 + (88)^2 + (220)^2}{4} - \frac{(808)^2}{16} = 5204.00$$

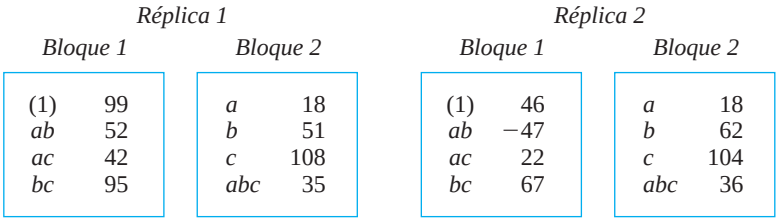


Figura 11.13 Datos para el ejemplo 11.15

Las demás sumas de cuadrados se calculan con el método de Yates utilizando los ocho totales de condición experimental y el resultado es la tabla ANOVA dada como tabla 11.14. Por comparación con $F_{.05,1,6} = 5.99$, se concluye que sólo los efectos principales para *A* y *C* difieren significativamente de cero.

Tabla 11.14 Tabla ANOVA para el ejemplo 11.15

Causa de la variación	Grados de libertad	Suma de cuadrados	Media cuadrática	*f*
A	1	12,996	12,996	39.82
B	1	702.25	702.25	2.15
C	1	2,756.25	2,756.25	8.45
AB	1	210.25	210.25	.64
AC	1	30.25	30.25	.093
BC	1	25	25	.077
Blocks	3	5,204	1,734.67	5.32
Error	6	1,958	326.33	
Total	15	23,882		

Confusión cuando se utilizan más de dos bloques

En el caso $r = 2$ (cuatro bloques), tres efectos se confunden con bloques. El experimentador primero selecciona dos efectos definitorios que han de ser confundidos. Por ejemplo, en un experimento con cinco factores (*A*, *B*, *C*, *D* y *E*), las dos interacciones de tres factores *BCD* y *CDE* podrían ser elegidas para confundirse. El tercer efecto confundido es entonces la **interacción generalizada** de los dos, obtenida escribiendo los dos efectos seleccionados uno al lado del otro y luego eliminando las letras cualesquiera comunes a ambos: $(BCD)(CDE) = BE$. Obsérvese que si *ABC* y *CDE* se eligen para confundirse, su interacción generalizada es $(ABC)(CDE) = ABDE$ de modo que ningunos efectos principales o interacciones de dos factores se confundan.

Una vez que los dos efectos definitorios hayan sido seleccionados para confundirse, un bloque se compone de todas las condiciones de tratamiento que tienen un número par de letras en común con ambos efectos definitorios. El segundo bloque se compone de todas las condiciones que tienen un número par de letras en común con el primer contraste definido y un número impar de letras en común con el segundo contraste y el tercero y cuarto bloques se componen de los contrastes “impar/par” e “impar/impar”. En un experimento con cinco factores con efectos definitorios *ABC* y *CDE*, esto da por resultado la asignación

de bloques como se muestra en la figura 11.14 (con el número de letras en común con cada contraste definitorio que aparece junto a cada condición experimental).

Bloque 1		*Bloque 2*		*Bloque 3*		*Bloque 4*	
(1)	(0, 0)	*d*	(0, 1)	*a*	(1, 0)	*c*	(1, 1)
ab	(2, 0)	*e*	(0, 1)	*b*	(1, 0)	*ad*	(1, 1)
de	(0, 2)	*ac*	(2, 1)	*cd*	(1, 2)	*ae*	(1, 1)
acd	(2, 2)	*bc*	(2, 1)	*ce*	(1, 2)	*bd*	(1, 1)
ace	(2, 2)	*abd*	(2, 1)	*ade*	(1, 2)	*be*	(1, 1)
bcd	(2, 2)	*abe*	(2, 1)	*bde*	(1, 2)	*abc*	(3, 1)
bce	(2, 2)	*acde*	(2, 3)	*abcd*	(3, 2)	*cde*	(1, 3)
abde	(2, 2)	*bcde*	(2, 3)	*abce*	(3, 2)	*abcde*	(3, 3)

Figura 11.14 Cuatro bloques en un experimento 2^5 factorial con efectos definitorios *ABC* y *CDE*

El bloque que contiene (1) se llama **bloque principal.** Una vez construido, se puede obtener un segundo bloque seleccionando cualquier condición experimental no incluida en el bloque principal y obteniendo su interacción generalizada con cada condición presente en el bloque principal. Se construyen entonces los demás bloques del mismo modo seleccionando primero una condición no incluida en un bloque ya construido y localizando interacciones generalizadas con el bloque principal.

En situaciones experimentales con $p > 3$, a menudo no existe ninguna réplica, así que las sumas de cuadrados asociadas con interacciones de alto grado no confundidas normalmente se agrupan para obtener una suma de cuadrados para error que pueda ser utilizada en el denominador de los varios estadísticos F. Todos los cálculos de nuevo se realizan con la técnica de Yates, con SSB1 como la suma de las sumas de cuadrados asociadas con efectos confundidos.

Cuando $r > 2$, primero se seleccionan r efectos definitorios que han de ser confundidos con bloques, asegurándose de que ninguno de los efectos elegidos sea la interacción generalizada de cualesquiera otros dos seleccionados. Los $2^r - r - 1$ efectos adicionales confundidos con los bloques son entonces interacciones generalizadas de todos los efectos presentes en el conjunto definitorio (incluidas no sólo las interacciones generalizadas de pares de efectos sino también conjuntos de tres, cuatro y así sucesivamente).

Réplica fraccionaria

Cuando el número p de factores es grande, incluso una sola réplica de un experimento 2^p puede ser cara y consumidora de tiempo. Por ejemplo, una réplica de un experimento 2^6 factorial implica una observación por cada una de las 64 condiciones experimentales diferentes. Una estrategia atractiva en tales situaciones es observar sólo una fracción de las 2^p condiciones. Siempre que se tenga cuidado en la elección de la condición que ha de ser observada, aún se puede obtener mucha información sobre efectos de factor.

Supóngase que se decide incluir sólo 2^{p-1} (la mitad) de las 2^p condiciones posibles en el experimento; esto normalmente se conoce como **media réplica.** El precio pagado por este ahorro es doble. Primero, la información sobre un solo efecto (determinada por las 2^{p-1} condiciones seleccionadas para observación) se pierde por completo para el experimentador en el sentido de que ninguna estimación razonable del efecto es posible. Segundo, los $2^p - 2$ efectos principales remanentes e interacciones se aparean de modo que cualquier efecto en un par particular se confunde con el otro efecto en el mismo par. Por ejemplo, un par como ése puede ser $\{A, BCD\}$, de modo que las estimaciones separadas del efecto principal A y de la interacción BCD no son posibles. Es deseable, entonces, seleccionar una media réplica con la cual los efectos principales y las interacciones de bajo grado sean apareadas (confundidas) sólo con interacciones de alto grado en lugar de una con otra.

El primer paso al especificar, una media réplica es seleccionar un efecto definitorio como el efecto no estimable. Supóngase que en un experimento con cinco factores, *ABCDE*

se elige como el efecto definitorio. Ahora las $2^5 = 32$ posibles condiciones de tratamiento se dividen en dos grupos con 16 condiciones cada uno, uno compuesto de todas las condiciones que tienen un número impar de letras en común con *ABCDE* y el otro que contiene un número par de letras en común con el contraste definido. Entonces cualquier grupo de 16 condiciones se utiliza como media réplica. El grupo "impar" es

$$a, b, c, d, e, abc, abd, abe, acd, ace, ade, bcd, bce, bde, cde, abcde$$

Cada efecto principal e interacción diferente de *ABCDE* se confunde entonces (**alias con)** con su interacción generalizada con *ABCDE*. Por lo tanto $(AB)(ABCDE) = CDE$, de tal suerte que la interacción *AB* y la interacción *CDE* se confunden entre sí. Los **pares de alias** resultantes son

$\{A, BCDE\}$	$\{B, ACDE\}$	$\{C, ABDE\}$	$\{D, ABCE\}$	$\{E, ABCD\}$
$\{AB, CDE\}$	$\{AC, BDE\}$	$\{AD, BCE\}$	$\{AE, BCD\}$	$\{BC, ADE\}$
$\{BD, ACE\}$	$\{BE, ACD\}$	$\{CD, ABE\}$	$\{CE, ABD\}$	$\{DE, ABC\}$

Obsérvese en particular que cada efecto principal se confunde con una interacción de cuatro factores. Suponiendo que estas interacciones son insignificantes se puede probar en cuanto a la presencia de efectos principales.

Para especificar un cuarto de réplica de un experimento 2^p factorial (2^{p-2} de las 2^p posibles condiciones de tratamiento), dos efectos definitorios deben ser seleccionados. Estos dos y su interacción generalizada se transforman en efectos no estimables. En lugar de pares de alias como en la media réplica, cada efecto restante ahora se confunde con otros tres efectos, y cada uno es su interacción generalizada con uno de los tres efectos no estimables.

Ejemplo 11.16 El artículo "More on Planning Experiments to Increase Research Efficiency" (*Industrial and Eng. Chemistry*, 1970: 60–65) reporta sobre los resultados de un cuarto de réplica de un experimento 2^5 en el cual los cinco factores fueron A = temperatura de condensación, B = cantidad de material B, C = volumen de solvente, D = tiempo de condensación y E = cantidad de material E. La variable de respuesta fue el rendimiento del proceso químico. Los contrastes definitorios seleccionados fueron *ACE* y *BDE*, con interacción generalizada $(ACE)(BDE) = ABCD$. Los 28 efectos principales e interacciones restantes ahora pueden ser divididos en siete grupos de cuatro efectos cada uno de modo que los efectos dentro de un grupo no pueden ser valorados por separado. Por ejemplo, las interacciones generalizadas de *A* con efectos no estimables son $(A)(ACE) = CE$, $(A)(BDE) = ABDE$ y $(A)(ABCD) = BCD$, de modo que un grupo alias es $\{A, CE, ABDE, BCD\}$. El conjunto completo de grupos alias es

$\{A, CE, ABDE, BCD\}$	$\{B, ABCE, DE, ACD\}$	$\{C, AE, BCDE, ABD\}$
$\{D, ACDE, BE, ABC\}$	$\{E, AC, BD, ABCDE\}$	$\{AB, BCE, ADE, CD\}$
$\{AD, CDE, ABE, BC\}$		

■

Una vez que se eligen los contrastes definitorios para un cuarto de réplica, se utilizan en la discusión de confusión para dividir las 2^p condiciones de tratamiento en cuatro grupos de 2^{p-2} condiciones cada uno. Entonces se selecciona uno de los cuatro grupos como el conjunto de condiciones en las cuales los datos serán recolectados. Comentarios similares aplican a una $\frac{1}{2}^r$ réplica de un experimento 2^p factorial.

Habiendo realizado observaciones para las combinaciones de tratamiento seleccionadas, se construye una tabla de signos similar a la tabla 11.10. La tabla contiene sólo un renglón por cada una de las combinaciones de tratamiento observadas en realidad en lugar de 2^p renglones y existe sólo una columna por cada grupo alias (puesto que cada efecto en el grupo tendría el mismo conjunto de signos en las condiciones de tratamiento seleccionadas para la observación). Los signos en cada columna indican como siempre cómo se calculan los contrastes para las diversas sumas de cuadrados. También se puede utilizar el método de Yates, pero la regla para disponer las condiciones observadas en orden estándar debe ser modificada.

La parte difícil del análisis de réplica fraccionaria en general implica decidir cuál utilizar para la suma de cuadrados de error. Puesto que normalmente no habrá réplica (aunque se podrían observar, por ejemplo, dos réplicas de un cuarto de réplica), algunas sumas de cuadrados para efectos deben ser agrupadas para obtener una suma de cuadrados para error. En una media réplica de un experimento 2^8, por ejemplo, se puede elegir una estructura alias de modo que cada uno de los ocho efectos principales y cada una de las 28 interacciones de dos factores se confundan sólo con interacciones de alto grado y que existan 27 grupos alias adicionales que impliquen sólo interacciones de alto grado. Suponiendo la ausencia de efectos de interacción de alto grado las 27 sumas de cuadrados resultantes pueden entonces ser sumadas para dar una suma de cuadrados para error, lo que permite pruebas de 1 grado de libertad de todos los efectos principales e interacciones de dos factores. Sin embargo, en muchos casos se pueden obtener pruebas de efectos principales sólo mediante la agrupación de algunas o todas las sumas de cuadrados asociadas con grupos alias que impliquen interacciones de dos factores y las correspondientes interacciones de dos factores no pueden ser investigadas.

Ejemplo 11.17 (Continuación del ejemplo 11.16)

El conjunto de condiciones de tratamiento seleccionadas y los resultados producidos para el cuarto de réplica del experimento 2^5 fueron

e	*ab*	*ad*	*bc*	*cd*	*ace*	*bde*	*abcde*
23.2	15.5	16.9	16.2	23.8	23.4	16.8	18.1

La tabla de signos abreviada se muestra en la tabla 11.15.

Con SSA denotando la suma de cuadrados para los efectos en el grupo alias {*A*, *CE*, *ABDE*, *BCD*}

$$\text{SSA} = \frac{(-23.2 + 15.5 + 16.9 - 16.2 - 23.8 + 23.4 - 16.8 + 18.1)^2}{8} = 4.65$$

Tabla 11.15 Tabla de signos para el ejemplo 11.17

	A	***B***	***C***	***D***	***E***	***AB***	***AD***
e	−	−	−	−	+	+	+
ab	+	+	−	−	−	+	−
ad	+	−	−	+	−	−	+
bc	−	+	+	−	−	−	+
cd	−	−	+	+	−	+	−
ace	+	−	+	−	+	−	−
bde	−	+	−	+	+	−	−
abcde	+	+	+	+	+	+	+

Asimismo SSB = 53.56, SSC = 10.35, SSD = .91, SSE′ = 10.35 (el ′ diferencía esta cantidad de la suma de cuadrados para error SSE), SSAB = 6.66 y SSAD = 3.25 y se obtiene SST = 4.65 + 53.56 + ⋯ + 3.25 = 89.73. Para probar en cuanto a efectos principales, se utiliza SSE = SSAB + SSAD = 9.91 con 2 grados de libertad. La tabla ANOVA aparece en la tabla 11.16.

Como $F_{.05,1,2} = 18.51$, ninguno de los cinco efectos principales puede ser juzgado como significativo. Desde luego, con sólo dos grados de libertad para error, la prueba no es muy poderosa (es decir, es bastante probable que no detecte la presencia de efectos). El artículo de *Industrial and Engineering Chemistry* de donde se tomaron los datos en realidad daba una estimación independiente del error estándar de los efectos de tratamiento basado en experiencias previas, de modo que utilizó un análisis algo diferente. El análisis aquí realizado fue sólo para propósitos ilustrativos, puesto que en general se desearía mucho más que 2 grados de libertad para error. ■

Tabla 11.16 Tabla ANOVA para el ejemplo 11.17

Causa	Grados de libertad	Suma de cuadrados	Media cuadrática	f
A	1	4.65	4.65	.94
B	1	53.56	53.56	10.80
C	1	10.35	10.35	2.09
D	1	.91	.91	.18
E	1	10.35	10.35	2.09
Error	2	9.91	4.96	
Total	7	89.73		

Como una alternativa de las pruebas F basadas en el agrupamiento de sumas de cuadrados para obtener SSE, se puede examinar una gráfica de probabilidad normal de contrastes de efecto.

Ejemplo 11.18 Se realizó un experimento para investigar la contracción de material plástico utilizado para fundas de cables de velocímetro ("An Explanation and Critique of Taguchi's Contribution to Quality Engineering", *Quality and Reliability Engr. Intl.*, 1988: 123–131). Los ingenieros comenzaron con 15 factores: diámetro externo del forro, troquel de forro, material de forro, velocidad de la línea de forrar, tipo de trenzado del alambre, tensión de trenzado, diámetro del alambre, tensión del forro, temperatura del forro, material de recubrimiento, tipo de troquel de recubrimiento, temperatura de fusión, paquete de pantalla, método de enfriamiento y velocidad de la línea. Se sospechaba que sólo algunos de estos factores eran importantes, así que se realizó un experimento de selección en la forma de un factorial 2^{15-11} (una $1/2^{11}$ fracción de un experimento 2^{15} factorial). La estructura alias resultante es bastante complicada; en particular, cada efecto principal se confunde con interacciones de dos factores. La variable de respuesta fue el porcentaje de contracción de un espécimen de cable producido a niveles diseñados de los factores.

La figura 11.15 muestra una gráfica de probabilidad normal de los contrastes de efecto. Todos excepto dos de los puntos se aproximan bastante a una línea recta. Los puntos discrepantes corresponden a los efectos E = tipo de trenzado del alambre y G = diámetro del alambre, lo que sugiere que esos dos factores son los únicos que afectan la cantidad de contracción.

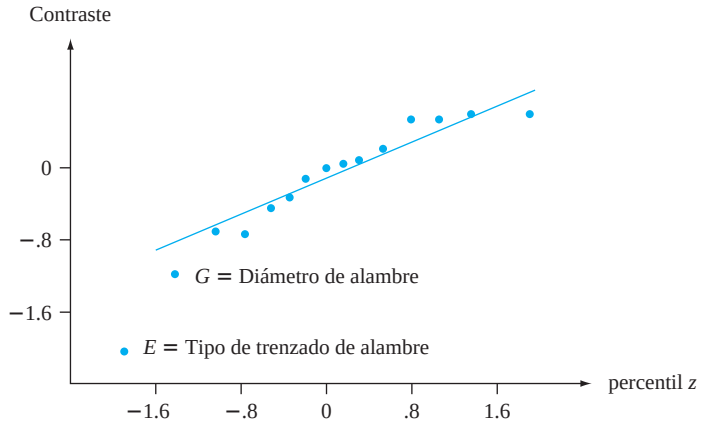


Figura 11.15 Gráfica de probabilidad normal de contrastes del ejemplo 11.18 ■

Los temas de la experimentación factorial, confusión y la replicación fraccionaria abarcan muchos modelos y técnicas que no hemos discutido. Por favor, consulte las referencias del capítulo para obtener más información.

EJERCICIOS Sección 11.4 (38–49)

38. Los datos adjuntos se obtuvieron con un experimento para estudiar la naturaleza de dependencia de la corriente de soldar en tres factores: voltaje de soldar, velocidad de alimentación del alambre y distancia de la punta del cautín a la pieza de trabajo. Hubo dos niveles de cada factor (un experimento 2^3) con dos réplicas por cada combinación de niveles (los promedios a través de las réplicas concuerdan con los valores dados en el artículo "A Study on Prediction of Welding Current in Gas Metal Arc Welding", *J. Engr. Manuf.*, 1991: 64–69). Los dos primeros números dados son para el tratamiento (1), los dos siguientes para *a* y así sucesivamente en orden estándar: 200.0, 204.2, 215.5, 219.5, 272.7, 276.9, 299.5, 302.7, 166.6, 172.6, 186.4, 192.0, 232.6, 240.8, 253.4, 261.6.

a. Verifique que las sumas de cuadrados son las que se dan en la tabla ANOVA adjunta generada por Minitab.

b. ¿Cuáles efectos parecen ser importantes y por qué?

Análisis de varianza para corriente

Fuente	DF	SS	MS	F	P
Voltaje	1	1685.1	1685.1	102.38	0.000
Velocidad	1	21272.2	21272.2	1292.37	0.000
Distancia	1	5076.6	5076.6	308.42	0.000
Voltaje*velocidad	1	36.6	36.6	2.22	0.174
Voltaje*distancia	1	0.4	0.4	0.03	0.877
Velocidad*distancia	1	109.2	109.2	6.63	0.033
Volt*vel*dist	1	23.5	23.5	1.43	0.266
Error	8	131.7	16.5		
Total	15	28335.3			

39. Los datos adjuntos se obtuvieron con un experimento 2^3 con tres réplicas por combinación de tratamientos diseñado para estudiar los efectos de concentración de detergente (*A*), concentración de carbonato de sodio (*B*) y concentración de celulosa carboximetilo de sodio (*C*) en el poder limpiador de una solución en pruebas de lavado (un número grande indica un mejor poder limpiador que uno pequeño):

Niveles de factor				
A	***B***	***C***	**Condición**	**Observaciones**
1	1	1	(1)	106, 93, 116
2	1	1	*a*	198, 200, 214
1	2	1	*b*	197, 202, 185
2	2	1	*ab*	329, 331, 307
1	1	2	*c*	149, 169, 135
2	1	2	*ac*	243, 247, 220
1	2	2	*bc*	255, 230, 252
2	2	2	*abc*	383, 360, 364

a. Tras de obtener los totales de celda x_{ijk}, calcule estimaciones de β_1, γ_{11}^{AC}, y γ_{21}^{AC}.

b. Use los totales de celda junto con el método de Yates para calcular los contrastes de efecto y las sumas de cuadrados. Luego construya una tabla ANOVA y pruebe todas las hipótesis apropiadas con $\alpha = .05$.

40. En un estudio de procesos utilizados para eliminar impurezas de artículos de celulosa ("Optimization of Rope-Range Bleaching of Cellulosic Fabrics", *Textile Research J.*, 1976: 493–496), se obtuvieron los siguientes datos con un experimento 2^4 que implica el proceso de desencolado. Los cuatro factores fueron concentración de enzima (*A*), pH (*B*), temperatura (*C*) y tiempo (*D*).

Tratamiento	Enzima (g/L)	pH	Temperatura (°C)	Tiempo (h)	% de almidón en peso 1a. répl.	% de almidón en peso 2a. réplica
(1)	.50	6.0	60.0	6	9.72	13.50
a	.75	6.0	60.0	6	9.80	14.04
b	.50	7.0	60.0	6	10.13	11.27
ab	.75	7.0	60.0	6	11.80	11.30
c	.50	6.0	70.0	6	12.70	11.37
ac	.75	6.0	70.0	6	11.96	12.05
bc	.50	7.0	70.0	6	11.38	9.92
abc	.75	7.0	70.0	6	11.80	11.10
d	.50	6.0	60.0	8	13.15	13.00
ad	.75	6.0	60.0	8	10.60	12.37
bd	.50	7.0	60.0	8	10.37	12.00
abd	.75	7.0	60.0	8	11.30	11.64
cd	.50	6.0	70.0	8	13.05	14.55
acd	.75	6.0	70.0	8	11.15	15.00
bcd	.50	7.0	70.0	8	12.70	14.10
abcd	.75	7.0	70.0	8	13.20	16.12

a. Use el algoritmo de Yates para obtener sumas de cuadrados y la tabla ANOVA.

b. ¿Parecen estar presentes efectos de interacción de segundo, tercero y cuarto orden? Explique su razonamiento. ¿Qué efectos principales parecen ser significativos?

41. En el ejercicio 39, suponga que se utilizó una baja temperatura del agua para obtener los datos. Se repite entonces todo el experimento con una temperatura del agua más alta para obtener los datos siguientes. Use el algoritmo de Yates con el conjunto completo de 48 observaciones para obtener las sumas de cuadrados y la tabla ANOVA y luego pruebe las hipótesis apropiadas al nivel .05.

Condición	Observaciones
d	144, 154, 158
ad	239, 227, 244
bd	232, 242, 246
abd	364, 362, 346
cd	194, 162, 203
acd	284, 295, 291
bcd	291, 287, 297
abcd	411, 406, 395

42. Los siguientes datos de consumo de energía en hornadas realizadas en un horno eléctrico (kW consumidos por tonelada de

producto fundido) se obtuvieron con un experimento factorial 2^4 con tres réplicas ("Studies on a 10-cwt Arc Furnace", *J. of the Iron and Steel Institute*, 1956: 22). Los factores fueron la naturaleza del techo *A* (bajo, alto), el ajuste de energía *B* (bajo, alto), chatarra utilizada *C* (tubo, placa) y carga *D* (700 lb, 1000 lb).

Tratamiento	x_{ijklm}	Tratamiento	x_{ijklm}
(1)	866, 862, 800	*d*	988, 808, 650
a	946, 800, 840	*ad*	966, 976, 876
b	774, 834, 746	*bd*	702, 658, 650
ab	709, 789, 646	*abd*	784, 700, 596
c	1017, 990, 954	*cd*	922, 808, 868
ac	1028, 906, 977	*acd*	1056, 870, 908
bc	817, 783, 771	*bcd*	798, 726, 700
abc	829, 806, 691	*abcd*	752, 714, 714

Construya la tabla ANOVA y pruebe todas las hipótesis de interés con $\alpha = .01$.

43. El artículo "Statistical Design and Analysis of Qualification Test Program for a Small Rocket Engine" (*Industrial Quality Control*, 1964: 14–18) presenta datos deducidos de un experimento para valuar los efectos de vibración (*A*), ciclaje de temperatura (*B*) ciclaje de altitud (*C*) y temperatura para ciclaje de altitud y encendido (*D*) sobre duración del empuje. Aquí se da un subconjunto de los datos. (En el artículo, hubo cuatro niveles de *D* en lugar de sólo dos.) Use el método de Yates para obtener sumas de cuadrados y la tabla ANOVA. Luego asuma que no existen interacciones de tres y cuatro factores, agrupe las sumas de cuadrados correspondientes para obtener una estimación de σ^2 y pruebe todas las hipótesis apropiadas al nivel .05.

		D_1		D_2	
		C_1	C_2	C_1	C_2
A_1	B_1	21.60	21.60	11.54	11.50
	B_2	21.09	22.17	11.14	11.32
A_2	B_1	21.60	21.86	11.75	9.82
	B_2	19.57	21.85	11.69	11.18

44. a. En un experimento 2^4, suponga que se van a utilizar dos bloques y que se decide confundir la interacción *ABCD* con el efecto de bloque. ¿Qué tratamientos deben ser realizados en el primer bloque [el que contiene el tratamiento (1)] y qué tratamientos se asignan al segundo bloque?

b. En un experimento para investigar la retención de niacina en vegetales en función de la temperatura de cocción (*A*), tamaño de cedazo (*B*), tipo de procesamiento (*C*) y tiempo de cocción (*D*), cada factor se mantuvo a dos niveles. Se utilizaron dos bloques, con la asignación de bloques como se da en el inciso (a) para confundir sólo la interacción *ABCD* con los bloques. Use el procedimiento de Yates para obtener la tabla ANOVA para los datos adjuntos.

Tratamiento	x_{ijkl}	Tratamiento	x_{ijkl}
(1)	91	*d*	72
a	85	*ad*	78
b	92	*bd*	68
ab	94	*abd*	79
c	86	*cd*	69
ac	83	*acd*	75
bc	85	*bcd*	72
abc	90	*abcd*	71

c. Suponga que no existen efectos tridireccionales, de modo que las sumas de cuadrados asociadas pueden ser combinadas para producir un estimado de σ^2 y realice todas las pruebas apropiadas al nivel .05.

45. a. Se realizó un experimento para investigar los efectos en la sensibilidad al audio de resistencia variable (*A*), dos capacitancias (*B*, *C*) e inductancia de una bobina (*D*) en una parte de un circuito de televisión. Si se utilizaron cuatro bloques con cuatro tratamientos por bloque y los efectos definitorios para confusión fueron *AB* y *CD*, ¿cuáles tratamientos aparecieron en cada bloque?

b. Suponga que se realizaron dos réplicas del experimento descrito en el inciso (a) y se obtuvieron los datos adjuntos. Obtenga la tabla ANOVA y pruebe todas las hipótesis pertinentes al nivel .01.

Tratamiento	x_{ijkl1}	x_{ijkl2}	Tratamiento	x_{ijkl1}	x_{ijkl2}
(1)	618	598	*d*	598	585
a	583	560	*ad*	587	541
b	477	525	*bd*	480	508
ab	421	462	*abd*	462	449
c	601	595	*cd*	603	577
ac	550	589	*acd*	571	552
bc	505	484	*bcd*	502	508
abc	452	451	*abcd*	449	455

46. En un experimento que implica cuatro factores (*A*, *B*, *C* y *D*) y cuatro bloques, demuestre que por lo menos un efecto principal o un efecto de interacción de dos factores debe ser confundido con el efecto de bloque.

47. a. En un experimento de siete factores (*A*, . . . , *G*), suponga que en realidad se realiza un cuarto de réplica. Si los efectos definitorios son *ABCDE* y *CDEFG*, ¿cuál es el tercer efecto no estimable y qué tratamientos están en el grupo que contiene (1)? ¿Cuáles son los grupos alias de los siete efectos principales?

b. Si el cuarto de réplica se tiene que realizar con cuatro bloques (con ocho tratamientos por bloque), ¿cuáles son los bloques si los efectos a ser confundidos son *ACF* y *BDG*?

48. El artículo "Applying Design of Experiments to Improve a Laser Welding Process" (*J. of Engr. Manufacture*, 2008: 1035–1042) incluye los resultados de una media réplica de un experimento 2^4. Los cuatro factores fueron: *A*. Potencia (2900 W, 3300 W), *B*. Corriente (2400 mV, 3600 mV), *C*. Limpieza de laterales (no, sí) y *D*. Limpieza de techo (no, sí).

a. Si el efecto *ABCD* es elegido como el efecto de definición para la réplica y el grupo de los ocho tratamientos para los que se obtienen los datos incluye el tratamiento (1), ¿qué otros tratamientos se observaron en el grupo y cuáles son los pares de alias?

b. El artículo citado presentó datos sobre dos variables de respuesta diferentes, el porcentaje de uniones defectuosas, para el cordón derecho de soldadura láser y el cordón de soldadura a la

izquierda. Aquí consideramos sólo la última respuesta. Las observaciones figuran en esta lista en orden estándar después de eliminar la mitad no observada. Suponiendo que interacciones de dos y tres factores son insignificantes, pruebe un nivel de .05 para la presencia de efectos principales. También construya una gráfica de probabilidad normal.

8.936 9.130 4.314 7.692
.415 6.061 1.984 3.830

49. Una media réplica de un experimento 2^5 para investigar el efecto del tiempo de tratamiento térmico (*A*), tiempo de temple (*B*), tiempo de estirado (*C*), posición de los serpentines calentadores (*D*) y posición de medición (*E*) en la dureza de piezas fundidas de acero, dio los datos adjuntos. Construya la tabla ANOVA y (suponiendo que interacciones de segundo orden y orden más alto son insignificantes) pruebe al nivel .01 en cuanto a la presencia de efectos principales. Construya también una gráfica de probabilidad normal.

Tratamiento	Observación	Tratamiento	Observación
a	70.4	*acd*	66.6
b	72.1	*ace*	67.5
c	70.4	*ade*	64.0
d	67.4	*bcd*	66.8
e	68.0	*bce*	70.3
abc	73.8	*bde*	67.9
abd	67.0	*cde*	65.9
abe	67.8	*abcde*	68.0

EJERCICIOS SUPLEMENTARIOS (50–61)

50. Los resultados de un estudio de la efectividad del secado en tendedero en la suavidad de telas se resumieron en el artículo "Line-Dried vs. Machine-Dried Fabrics: Comparison of Appearance, Hand, and Consumer Acceptance" (*Home Econ. Research J.*, 1984: 27–35). Se dieron calificaciones de suavidad para nueve tipos diferentes de telas y cinco métodos de secado diferentes: (1) secado en máquina, (2) secado colgado, (3) secado colgado seguido por 15 min de secado en centrífuga, (4) secado colgado con suavizante y (5) secado colgado con movimiento de aire. Considerando los diferentes tipos de tela como bloques, construya una tabla ANOVA. Con un nivel de significancia de .05, pruebe para ver si hay diferencia entre la calificación de suavidad media real según los métodos de secado.

		Método de secado				
		1	2	3	4	5
Tela	Crepé	3.3	2.5	2.8	2.5	1.9
	Doble punto	3.6	2.0	3.6	2.4	2.3
	Asargada	4.2	3.4	3.8	3.1	3.1
	Asargada combinada	3.4	2.4	2.9	1.6	1.7
	Tela esponja	3.8	1.3	2.8	2.0	1.6
	Paño fino	2.2	1.5	2.7	1.5	1.9
	Lencería para sábanas	3.5	2.1	2.8	2.1	2.2
	Pana	3.6	1.3	2.8	1.7	1.8
	Mezclilla	2.6	1.4	2.4	1.3	1.6

51. La absorción de agua de dos tipos de mortero utilizado para reparar cemento dañado se discutió en el artículo "Polymer Mortar Composite Matrices for Maintenance-Free, Highly Durable Ferrocement" (*J. of Ferrocement*, 1984: 337–345). Se sumergieron especímenes de mortero para cemento común (OCM, por sus siglas en inglés) y mortero para cemento con polímero (PCM) durante lapsos de tiempo variables (5, 9, 24 o 48 horas) y se registró la absorción de agua (% por peso). Con el tipo de mortero como factor *A* (con dos niveles) y el periodo de inmersión como factor *B* (con cuatro niveles), se realizaron tres observaciones por cada combinación de nivel de factor. Se utilizaron datos incluidos en el artículo para calcular las sumas de cuadrados, las cuales fueron SSA = 322.667, SSB = 35.623, SSAB = 8.557, y SST = 372.113. Use esta información para construir una tabla ANOVA. Pruebe las hipótesis apropiadas a un nivel de significancia de .05.

52. Se dispuso de cuatro parcelas para un experimento para comparar la acumulación de tréboles con cuatro tipos de sembrado ("Performance of Overdrilled Red Clover with Different Sowing Rates and Initial Grazing Managements", *N. Zeal. J. of Exp. Ag.*, 1984: 71–81). Como las cuatro parcelas habían sido rozadas de forma diferente antes del experimento y se pensaba que esto podía afectar la acumulación de tréboles, se utilizó un experimento de bloques aleatorizados con los cuatro tipos de sembrado probados en una sección de cada parcela. Use los datos dados para probar la hipótesis nula de ninguna diferencia en la acumulación de trébol media verdadera (kg DM/ha) con los diferentes tipos de sembrado.

		Coeficiente de sembrado (kg/ha)			
		3.6	6.6	10.2	13.5
Parcela	1	1155	2255	3505	4632
	2	123	406	564	416
	3	68	416	662	379
	4	62	75	362	564

53. En un proceso de control químico automatizado, la velocidad con la cual los objetos colocados sobre una banda transportadora pasan a través de un rociado químico (velocidad de banda), la cantidad de químico rociado (volumen rociado) y la marca del químico utilizado (marca) son factores que pueden afectar la uniformidad del recubrimiento aplicado. Se condujo

un experimento 2^3 replicado en un esfuerzo por incrementar la uniformidad del recubrimiento. En la tabla siguiente, los valores altos de la variable de respuesta están asociados con la alta uniformidad superficial:

Corrida	Volumen de rocío	Velocidad de banda	Marca	Uniformidad superficial Réplica 1	Uniformidad superficial Réplica 2
1	−	−	−	40	36
2	+	−	−	25	28
3	−	+	−	30	32
4	+	+	−	50	48
5	−	−	+	45	43
6	+	−	+	25	30
7	−	+	+	30	29
8	+	+	+	52	49

Analice estos datos y exponga sus conclusiones.

54. Plantas de energía de carbón utilizadas en la industria eléctrica han captado la atención del público debido a los problemas ambientales asociados con los desechos sólidos generados por la combustión a gran escala (“Fly Ash Binders in Stabilization of FGD Wastes”, *J. of Environmental Engineering*, 1998: 43–49). Se realizó un estudio para analizar la influencia de tres factores: tipo de aglutinante (*A*), cantidad de agua (*B*) y escenario de disposición en la tierra (*C*) que afectan ciertas características de lixiviación de los desechos sólidos resultantes de la combustión. Cada factor se estudió a dos niveles. Se realizó un experimento 2^3 no replicado y se midió un valor de respuesta EC50 (la concentración efectiva, en mg/L, que reduce el 50% de la luz en un bioensayo de luminiscencia) por cada combinación de niveles de factor. Los datos experimentales se dan en la siguiente tabla:

Corrida	Factor *A*	Factor B	Factor *C*	Respuesta EC50
1	−1	−1	−1	23,100
2	1	−1	−1	43,000
3	−1	1	−1	71,400
4	1	1	−1	76,000
5	−1	−1	1	37,000
6	1	−1	1	33,200
7	−1	1	1	17,000
8	1	1	1	16,500

Lleve a cabo un ANOVA apropiado y establezca sus conclusiones.

55. Las impurezas en la forma de óxidos de hierro reducen el valor económico y la utilidad de minerales industriales, tales como caolines, en las industrias de la cerámica y de procesamiento de papel. Se realizó un experimento 2^4 para valuar los efectos de cuatro factores en el porcentaje de hierro extraído de muestras de caolín (“Factorial Experiments in the Development of a Kaolin Bleaching Process Using Thiourea in Sulphuric Acid Solutions”, *Hydrometallurgy*, 1997: 181–197). Los factores y sus niveles aparecen en la siguiente tabla:

Factor	Descripción	Unidades	Nivel bajo	Nivel alto
A	H_2SO_4	M	.10	.25
B	Tiourea	g/L	0.0	5.0
C	Temperatura	°C	70	90
D	Tiempo	min	30	150

Los datos obtenidos de un experimento 2^4 no replicado se dan en la tabla siguiente.

Hornada de prueba	Extracción de hierro (%)	Hornada de prueba	Extracción de hierro (%)
(1)	7	*d*	28
a	11	*ad*	51
b	7	*bd*	33
ab	12	*abd*	57
c	21	*cd*	70
ac	41	*acd*	95
bc	27	*bcd*	77
abc	48	*abcd*	99

a. Calcule estimaciones de los efectos principales y los efectos de interacción de dos factores para este experimento.
b. Cree una gráfica de probabilidad de los efectos. ¿Cuáles efectos parecen ser importantes?

56. Se han utilizado diseños factoriales en la silvicultura para valuar los efectos de varios factores en el comportamiento de crecimiento de árboles. En el experimento, los investigadores pensaban que los retoños de abetos sanos debían abotonar más pronto que los retoños de abetos enfermos (“Practical Analysis of Factorial Experiments in Forestry”, *Canadian J. of Forestry*, 1995: 446–461). Además, antes de plantarlos, los retoños fueron expuestos a tres niveles de pH para ver si este factor tenía algún efecto en la captación de virus por las raíces. La tabla siguiente muestra datos de un experimento de 2×3 para estudiar ambos factores.

		pH 3	pH 5.5	pH 7
	Enfermo	1.2, 1.4, 1.0, 1.2, 1.4	.8, .6, .8, 1.0, .8	1.0, 1.0, 1.2, 1.4, 1.2
Salud	**Saludable**	1.4, 1.6, 1.6, 1.6, 1.4	1.0, 1.2, 1.2, 1.4, 1.4	1.2, 1.4, 1.2, 1.2, 1.4

La variable de respuesta es una calificación promedio de cinco botones de un retoño. Las calificaciones son 0 (botón no abierto) 1 (botón parcialmente abierto) y 2 (botón totalmente abierto). Analice estos datos.

57. Una propiedad de las bolsas de aire automotrices que contribuye a su capacidad de absorber energía es la permeabilidad ($pie^3/pie^2/min$) del material tejido utilizado para construir las bolsas de aire. Entender cómo la permeabilidad es influenciada

por varios factores es importante para incrementar la efectividad de las bolsas de aire. En un estudio, se analizaron los efectos de tres factores, cada uno a tres niveles. ("Analysis of Fabrics Used in Passive Restraint Systems—Airbags", *J. of the Textile Institute*, 1996: 554–571):

A (Temperatura): 8°C, 50°C, 75°C

B (Denier de la tela): 420-D, 630-D, 840-D

C (Presión del aire): 17.2 kPa, 34.4 kPa, 103.4 kPa

Temperatura 8°

	Presión		
Denier	**17.2**	**34.4**	**103.4**
420-D	73	157	332
	80	155	322
630-D	35	91	288
	433	98	271
840-D	125	234	477
	111	233	464

Temperatura 50°

	Presión		
Denier	**17.2**	**34.4**	**103.4**
420-D	52	125	281
	51	118	264
630-D	16	72	169
	12	78	173
840-D	96	149	338
	100	155	350

Temperatura 75°

	Presión		
Denier	**17.2**	**34.4**	**103.4**
420-D	37	95	276
	31	106	281
630-D	30	91	213
	41	100	211
840-D	102	170	307
	98	160	311

Analice estos datos y exprese sus conclusiones (suponga que todos los factores son fijos).

58. Un ingeniero químico ha realizado un experimento para estudiar los efectos de los factores fijos de presión de cuba (*A*), tiempo de cocción de la pulpa (*B*) y concentración de madera dura (*C*) en la resistencia del papel. El experimento implicó dos presiones, cuatro tiempos de cocción, tres concentraciones y dos observaciones con cada combinación de estos niveles. Las sumas calculadas de los cuadrados son SSA = 6.94, SSB = 5.61, SSC = 12.33, SSAB = 4.05, SSAC = 7.32, SSBC = 15.80, SSE = 14.40 y SST = 70.82. Construya la tabla ANOVA y realice pruebas apropiadas a un nivel de significancia de .05.

59. Se estudió la resistencia de adhesión cuando se monta un circuito integrado en un sustrato de vidrio metalizado como una función del factor *A* = tipo de adhesivo, factor *B* = curva tiempo y factor *C* = material conductor (cobre y níquel). Los datos se dan a continuación junto con una tabla ANOVA generada por Minitab. ¿Qué conclusiones puede sacar de los datos?

		Tiempo de curado		
Cobre		**1**	**2**	**3**
		72.7	74.6	80.0
	1	80.0	77.5	82.7
		77.8	78.5	84.6
Adhesivo	**2**	75.3	81.1	78.3
		77.3	80.9	83.9
	3	76.5	82.6	85.0
Níquel		**1**	**2**	**3**
		74.7	75.7	77.2
	1	77.4	78.2	74.6
		79.3	78.8	83.0
Adhesivo	**2**	77.8	75.4	83.9
		77.2	84.5	89.4
	3	78.4	77.5	81.2

```
Análisis de varianza para resistencia
Fuente                GL       SS       MS      F      P
Adhesivo               2  101.317   50.659   6.54  0.007
Tiempo de curado       2  151.317   75.659   9.76  0.001
Matcond                1    0.722    0.722   0.09  0.764
Adhesivo*
tiempo de curado       4   30.526    7.632   0.98  0.441
Adhesivo*matcond       2    8.015    4.008   0.52  0.605
Tiempo de curado*
matcond                2    5.952    2.976   0.38  0.687
Adhesivo*tiempo
decurado*matcond       4   33.298    8.325   1.07  0.398
Error                 18  139.515    7.751
Total                 35  470.663
```

60. El artículo "Effect of Cutting Conditions on Tool Performance in CBN Hard Turning" (*J. of Manuf. Processes*, 2005: 10–17) reportó los datos adjuntos de la velocidad de corte (m/s), alimentación (mm/rev), profundidad de corte (mm) y la vida de la herramienta (min). Lleve a cabo un análisis de varianza de tres factores de la vida de la herramienta, asumiendo la ausencia de cualquier posible interacción de los factores (como lo hicieron los autores del artículo).

Observación	Velocidad de corte	Alimentación	Profundidad de corte	Vida
1	1.21	0.061	0.102	27.5
2	1.21	0.168	0.102	26.5
3	1.21	0.061	0.203	27.0
4	1.21	0.168	0.203	25.0
5	3.05	0.061	0.102	8.0
6	3.05	0.168	0.102	5.0
7	3.05	0.061	0.203	7.0
8	3.05	0.168	0.203	3.5

61. Análogo a un cuadrado latino, se puede utilizar un diseño de cuadrado grecolatino cuando se sospecha que tres factores

externos pueden afectar las variables de respuesta y los cuatro factores (los tres externos y el de interés) tienen el mismo número de niveles. En un cuadrado latino, cada nivel del factor de interés (C) aparece una vez en cada renglón (con cada nivel de A) y una vez en cada columna (con cada nivel de B). En un cuadrado greco-latino, cada nivel del factor D aparece una vez en cada renglón, en cada columna y también con cada nivel del tercer factor externo C. Alternativamente, se puede utilizar el diseño cuando los cuatro factores son de igual interés, el número de niveles de cada uno es N y están disponibles recursos para sólo N^2 observaciones. Un cuadrado de 5×5 se ilustra en (a) con (k, l) en cada celda que denota el k-ésimo nivel de C y el l-ésimo nivel de D. En (b) se presentan datos de pérdida de peso en barras de silicio utilizadas para material semiconductor como una función del volumen de grabado al agua fuerte (A), color del ácido nítrico en solución de grabado (B), tamaño de las barras (C) y tiempo en la solución de grabado (D) (de "Applications of Analytic Techniques to the Semiconductor Industry", Fourteenth Midwest Quality Control Conference, 1959).

Sea $x_{ij(kl)}$ la pérdida de peso observada cuando el factor A está al nivel i, B está al nivel j, C está al nivel k y D está al nivel l. Suponiendo que no hay interacción entre los factores, la suma total de cuadrados SST (con $N^2 - 1$ grados de libertad) puede ser dividida en SSA, SSB, SSC, SSD y SSE. Dé expresiones para estas sumas de cuadrados, incluidas las fórmulas, obtenga la tabla ANOVA de los datos dados y pruebe cada una de las cuatro hipótesis de efecto principal con $\alpha = .05$.

		B				
(C, D)	**1**	**2**	**3**	**4**	**5**	
	1	(1, 1)	(2, 3)	(3, 5)	(4, 2)	(5, 4)
	2	(2, 2)	(3, 4)	(4, 1)	(5, 3)	(1, 5)
A	**3**	(3, 3)	(4, 5)	(5, 2)	(1, 4)	(2, 1)
	4	(4, 4)	(5, 1)	(1, 3)	(2, 5)	(3, 2)
	5	(5, 5)	(1, 2)	(2, 4)	(3, 1)	(4, 3)

(a)

65	82	108	101	126
84	109	73	97	83
105	129	89	89	52
119	72	76	117	84
97	59	94	78	106

(b)

Bibliografía

Box, George, William Hunter y Stuart Hunter, *Statistics for Experimenters* (2a. ed.), Wiley, Nueva York, 2006. Contiene un caudal de sugerencias e ideas sobre análisis de datos basados en la extensa experiencia consultora de los autores.

DeVor, R., T. Chang y J. W. Sutherland, *Statistical Quality Design and Control*, (2a. ed.), Prentice-Hall, Englewood Cliffs, NJ, 2006. Incluye un estudio moderno de experimentos factoriales y factoriales fraccionarios con un mínimo de matemáticas.

Hocking, Ronald, *Methods and Applications of Linear Models*, (2a. ed.), Wiley, Nueva York, 2003. Un tratamiento muy general de análisis de varianza escrito por una de las autoridades más reconocidas en este campo.

Kleinbaum, David, Lawrence Kupper, Keith Muller y Azhar Nizam, *Applied Regression Analysis and Other Multivariable Methods* (4th. ed.) Duxbury Press, Boston, 2007. Contiene una discusión especialmente buena de problemas asociados con el análisis de "datos desbalanceados", es decir K_{ij} desiguales.

Kuehl, Robert O., *Design of Experiments: Statistical Principles of Research Design and Analysis*, (2a. ed.), Duxbury Press, Boston, 1999. Un tratamiento amplio y actualizado de experimentos diseñados y análisis de los datos resultantes.

Montgomery, Douglas, *Design and Analysis of Experiments* (7a. ed.), Wiley, Nueva York, 2009. Véase la bibliografía del capítulo 10.

Neter, John, William Wasserman y Michael Kutner, *Applied Linear Statistical Models* (5a. ed.), Irwin, Homewood, IL., 2004. Véase la bibliografía del capítulo 10.

Vardeman, Stephen, *Statistics for Engineering Problem Solving*, PWS, Boston, 1994. Una introducción general para ingenieros, con mucha metodología descriptiva e inferencial para datos obtenidos de experimentos diseñados.

12 Regresión lineal simple y correlación

INTRODUCCIÓN

En los problemas de dos muestras discutidos en el capítulo 9, interesaba comparar valores de parámetros para la distribución x y la distribución y. Incluso cuando las observaciones se aparearon, no se trató de utilizar información sobre una de las variables al estudiar la otra variable. Éste es precisamente el objetivo del análisis de regresión, explotar la relación entre dos (o más) variables de modo que se pueda obtener información sobre una de ellas mediante el conocimiento de los valores de la otra u otras.

Una gran parte de las matemáticas está dedicada a estudiar variables que están *determinísticamente* relacionadas. Decir que x y y están relacionadas de esta manera significa que una vez se conoce el valor de x, el valor de y queda completamente especificado. Por ejemplo, supóngase que se decide rentar una vagoneta por un día y la renta es de \$25.00 más \$.30 por milla recorrida. Si x = el número de millas recorridas y y = el cargo de la renta, entonces $y = 25 + .3x$. Si se recorren 100 millas ($x = 100$), entonces $x = 25 + .3(100) = 55$. Como otro ejemplo, si la velocidad inicial de una partícula es v_0 y experimenta una aceleración constante a, entonces la distancia recorrida $= y = v_0x + \frac{1}{2}ax^2$, donde x = tiempo.

Existen muchas variables x y y que parecerían estar relacionadas entre sí, pero no de una forma determinística. Un ejemplo conocido para muchos estudiantes está dado por las variables x = promedio de calificaciones de preparatoria (GPA, por sus siglas en inglés) y y = GPA universitario. El valor de y no puede ser determinado sólo con el conocimiento de x y dos estudiantes diferentes podrían tener el mismo valor x pero tener valores y muy diferentes. No obstante existe la tendencia de que aquellos estudiantes que tienen promedio de calificaciones de preparatoria alto (bajo) también tienen un promedio de calificaciones universitario alto (bajo). El conocimiento del promedio de calificaciones de preparatoria de un estudiante es bastante útil ya que permite predecir cómo se desempeñará esa persona en la universidad.

Otros ejemplos de variables relacionadas de una forma no determinística incluyen x = edad de un niño y y = tamaño del vocabulario de ese niño, x = tamaño de un motor en centímetros cúbicos y y = eficiencia de combustible de un automóvil equipado con dicho motor y x = fuerza de tensión aplicada y y = cantidad de alargamiento en una tira de metal.

El **análisis de regresión** es la parte de la estadística que se ocupa de investigar la relación entre dos o más variables relacionadas en una forma no determinística. En este capítulo, se generaliza la relación lineal determinística $y = \beta_0 + \beta_1 x$ a una relación probabilística lineal, se desarrollan procedimientos para hacer inferencias sobre del modelo y se obtiene una medida cuantitativa (el coeficiente de correlación) del grado al cual las dos variables están relacionadas. En el siguiente capítulo, se considerarán técnicas para validar un modelo particular y para investigar relaciones no lineales y relaciones que implican más de dos variables.

12.1 Modelo de regresión lineal simple

La relación matemática determinística más simple entre dos variables x y y es una relación lineal $y = \beta_0 + \beta_1 x$. El conjunto de pares (x, y) para los cuales $y = \beta_0 + \beta_1 x$ determina una línea recta con pendiente β_1 e intersección en y β_0.* El objetivo de esta sección es desarrollar un modelo probabilístico lineal.

Si las dos variables no están determinísticamente relacionadas, entonces con un valor fijo de x, el valor de la segunda variable es incierto. Por ejemplo, si se está investigando la relación entre la edad de un niño y el tamaño del vocabulario y se decide seleccionar un niño de $x = 5.0$ años de edad, entonces antes de hacer la selección, el tamaño del vocabulario es una variable aleatoria Y. Después de que un niño particular de 5 años de edad ha sido seleccionado y sometido a prueba, el resultado puede ser un vocabulario de 2000 palabras. Se diría entonces que el valor observado de Y asociado con la fijación de $x = 5.0$ fue $y = 2000$.

Más generalmente, la variable cuyo valor es fijado por el experimentador será denotada por x y se llamará **variable independiente, pronosticadora o explicativa**. Con x fija, la segunda variable será aleatoria; esta variable aleatoria y su valor observado se designan por Y y y, respectivamente y se la conoce como **variable dependiente** o de **respuesta.**

Normalmente se realizarán observaciones para varios escenarios de la variable independiente. Sean $x_1, x_2, \ldots, x_n$ los valores de la variable independiente para la que se realizan las observaciones y sean Y_i y y_i respectivamente la variable aleatoria y el valor observado asociado con x: Los datos bivariantes disponibles se componen entonces de los n pares $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. Una imagen de estos datos llamada gráfica de dispersión proporciona impresiones preliminares acerca de la naturaleza de cualquier relación. En una gráfica como esa, cada (x_i, y_i) está representado como un punto colocado en un sistema de coordenadas bidimensional.

* La pendiente de una recta es el cambio en y con un incremento de 1 unidad en x. Por ejemplo, si $y = -3x + 10$, entonces y se reduce en 3 cuando x se incrementa en 1, de modo que la pendiente es -3. La intersección en y es la altura a la cual la línea cruza el eje vertical y se obtiene haciendo $x = 0$ en la ecuación.

Ejemplo 12.1 Los problemas visuales y musculoesqueléticos asociados con el uso de terminales con pantalla de visualización (VDT, por sus siglas en inglés) se han vuelto un tanto comunes en años recientes. Algunos investigadores se han enfocado en la dirección vertical de la mirada fija como causa del cansancio e irritación de los ojos. Se sabe que esta relación está estrechamente relacionada con el área de la superficie ocular (OSA, por sus siglas en inglés), así que se requiere un método de medir el área de la superficie ocular. Los datos representativos adjuntos sobre y = OSA (cm^2) y x = ancho de la fisura palpebral (es decir, el ancho horizontal de la apertura del ojo, en cm) se tomó del artículo "Analysis of Ocular Surface Area for Comfortable VDT Workstation Layout" (*Ergonomics*, 1996: 877–884). No se da el orden en el cual se obtuvieron las observaciones, así que por conveniencia aparecen en orden creciente de los valores x.

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_i	.40	.42	.48	.51	.57	.60	.70	.75	.75	.78	.84	.95	.99	1.03	1.12
y_i	1.02	1.21	.88	.98	1.52	1.83	1.50	1.80	1.74	1.63	2.00	2.80	2.48	2.47	3.05

i	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
x_i	1.15	1.20	1.25	1.25	1.28	1.30	1.34	1.37	1.40	1.43	1.46	1.49	1.55	1.58	1.60
y_i	3.18	3.76	3.68	3.82	3.21	4.27	3.12	3.99	3.75	4.10	4.18	3.77	4.34	4.21	4.92

Por consiguiente, $(x_1, y_1) = (.40, 1.02)$, $(x_5, y_5) = (.57, 1.52)$, y así sucesivamente. En la figura 12.1 se muestra una gráfica de dispersión obtenida con Minitab; se utilizó una opción que produjo una gráfica de puntos tanto de valores x como de valores y individualmente a lo largo de los márgenes derecho y superior de la gráfica, lo que facilita visualizar las distribuciones de las variables individuales (los histogramas o gráficas de caja son opciones alternativas). He aquí algunas cosas que hay que tener en cuenta sobre los datos y la gráfica:

- Varias observaciones tienen valores x idénticos aunque valores y diferentes (p. ej., $x_8 = x_9 = .75$ pero $y_8 = 1.80$ y $y_9 = 1.74$). Por lo tanto el valor de y *no* está determinado sólo por x sino también por varios otros factores.
- Existe una fuerte tendencia de que y se incremente a medida que x lo hace. Es decir, los valores grandes de OSA tienden a asociarse con valores grandes de ancho de fisura, una relación positiva entre las variables.

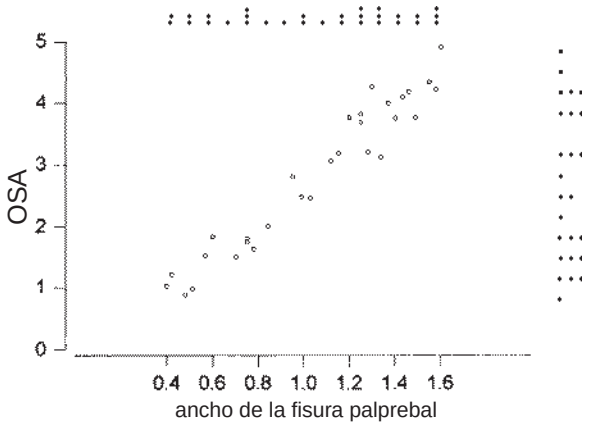


Figura 12.1 Gráfica de dispersión obtenida con Minitab con los datos del ejemplo 12.1, junto con gráficas de puntos de valores x y y

- Parece que el valor de y podría ser pronosticado a partir de x encontrando una línea que esté razonablemente cerca a los puntos presentes en la gráfica (los autores del artículo citado superponen tal línea en su gráfica). En otras palabras, existe evidencia de una relación lineal (aunque no perfecta) sustancial entre las dos variables. ■

Los ejes horizontal y vertical en la gráfica de dispersión de la figura 12.1 se cortan en el punto (0, 0). En muchos conjuntos de datos de x o y de los valores de ambas variables difieren considerablemente de cero con respecto al rango o rangos de los valores. Por ejemplo, un estudio de cómo la eficiencia de un equipo de aire acondicionado está relacionada con la temperatura diaria máxima a la intemperie podría implicar observaciones de temperaturas desde 80°F hasta 100°F. Cuando éste es el caso, una gráfica más informativa mostraría los ejes apropiadamente marcados intersectándose en algún punto diferente de (0, 0).

Ejemplo 12.2 El arsénico se encuentra en muchas aguas subterráneas y algunas aguas superficiales. Investigaciones recientes sobre sus efectos en la salud ha llevado a la Agencia de Protección Ambiental a reducir los niveles permisibles de arsénico en el agua potable de manera que muchos sistemas de agua ya no son compatibles con las normas. Este interés ha estimulado el desarrollo de métodos para eliminar el arsénico. Los datos adjuntos en x = pH y y = arsénico eliminado (%) por un proceso en particular fue leído de un gráfico de dispersión en el artículo "Optimizing Arsenic Removal During Iron Removal: Theoretical and Practical Considerations" (*J. of Water Supply Res. and Tech.*, 2005: 545–560).

x	7.01	7.11	7.12	7.24	7.94	7.94	8.04	8.05	8.07
y	60	67	66	52	50	45	52	48	40
x	8.90	8.94	8.95	8.97	8.98	9.85	9.86	9.86	9.87
y	23	20	40	31	26	9	22	13	7

La figura 12.2 muestra dos gráficas de dispersión de estos datos obtenidas con Minitab. En la figura 12.2(a), Minitab seleccionó la escala para ambos ejes. La figura 12.2(b) se obtuvo especificando una escala para los ejes de modo que se intersectan aproximadamente en el punto (0, 0). La segunda gráfica está más amontonada que la primera; tal amontonamiento hace más difícil valorar la naturaleza general de cualquier relación. Por ejemplo, puede ser más difícil descubrir la curvatura en una gráfica amontonada.

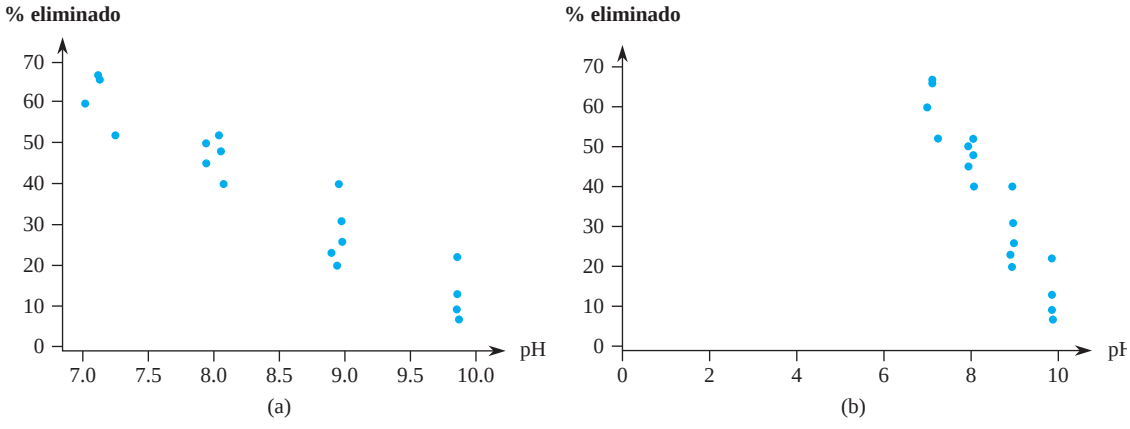


Figura 12.2 Gráficas de dispersión obtenidas con Minitab con los datos del ejemplo 12.2

Los grandes valores de arsénico eliminado tienden a asociarse con un bajo pH, una relación negativa o inversa. Además, las dos variables parecen estar al menos aproximadamente relacionadas de forma lineal, aunque los puntos en la gráfica se dispersarían en torno a cualquier línea recta sobrepuesta (tal recta aparece en la gráfica en el artículo citado). ■

Modelo probabilístico lineal

Para el modelo determinístico $y = \beta_0 + \beta_1 x$, el valor observado real de y es una función lineal de x. La generalización apropiada de esto a un modelo probabilístico asume que *el valor esperado de Y es una función lineal de x*, pero que con x fija, la variable Y difiere de su valor esperado en una cantidad aleatoria.

DEFINICIÓN

Modelo de regresión lineal simple

Existen parámetros β_0, β_1 y σ^2 de tal suerte que con cualquier valor fijo de la variable independiente x, la variable dependiente es una variable aleatoria y está relacionada con x por conducto de la **ecuación de modelo**

$$Y = \beta_0 + \beta_1 x + \epsilon \qquad (12.1)$$

La cantidad ϵ en la ecuación de modelo es una variable aleatoria, que se supone está normalmente distribuida con $E(\epsilon) = 0$ y $V(\epsilon) = \sigma^2$.

La variable ϵ se conoce como **término de error aleatorio** o **desviación aleatoria** en el modelo. Sin ϵ, cualquier par observado (x, y) correspondería a un punto que queda exactamente sobre la línea $y = \beta_0 + \beta_1 x$, llamada **línea de regresión** (o de **población**) **verdadera**. La inclusión del término de error aleatorio permite a (x, y) quedar por encima de la línea de regresión (cuando $\epsilon > 0$ o por debajo (cuando $\epsilon < 0$. Los puntos $(x_1, y_1), \ldots, (x_n, y_n)$ provenientes de n observaciones independientes se dispersarán entonces en torno a la línea de regresión verdadera, como se ilustra en la figura 12.3. En ocasiones, la conveniencia del modelo de regresión lineal simple puede ser sugerida por consideraciones teóricas (p. ej., existe una relación lineal exacta entre las dos variables, con ϵ representando el error de medición). Con mucha más frecuencia, no obstante, la racionalidad del modelo es indicada por una gráfica de dispersión que exhibe un patrón lineal sustancial (como en las figuras 12.1 y 12.2).

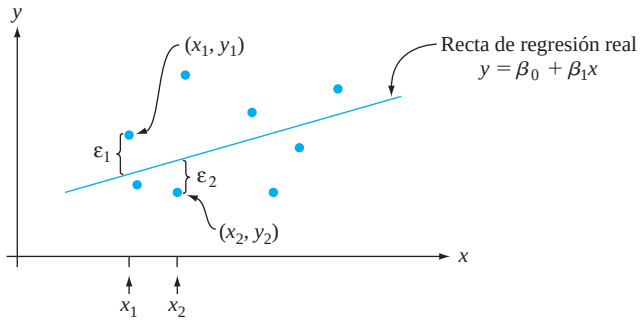


Figura 12.3 Puntos correspondientes a observaciones del modelo de regresión lineal simple

Las implicaciones de la ecuación de modelo (12.1) se entienden mejor con la ayuda de la siguiente notación. Sea x^* un valor particular de la variable independientes x y

$$\mu_{Y\cdot x^*} = \text{el valor esperado (o media) de } Y \text{ cuando } x = x^*$$
$$\sigma^2_{Y\cdot x^*} = \text{la varianza de } Y \text{ cuando } x = x^*$$

La notación alternativa es $E(Y|x^*)$ y $V(Y|x^*)$. Por ejemplo, si x = esfuerzo aplicado $(\text{kg/mm})^2$ y y = tiempo para la fractura (h), entonces $\mu_{Y\cdot 20}$ denotaría el valor esperado de tiempo para la fractura cuando se aplica un esfuerzo de 20 kg/mm^2. Si se piensa en una población completa de pares (x, y), entonces $\mu_{Y\cdot x^*}$ es la media de todos los valores y con los cuales $x = x^*$ y $\sigma^2_{Y\cdot x^*}$ es una medida de cuántos de estos valores de y se dispersan en torno al valor medio. Si por ejemplo, x = edad de un niño y y = tamaño del vocabulario, entonces $\mu_{Y\cdot 5}$ es el tamaño de vocabulario promedio de todos los niños de 5 años que hay en la población y $\sigma^2_{Y\cdot 5}$ describe la cantidad de variabilidad del tamaño de vocabulario de esta parte de la población. Una vez que se fija x, la única aleatoriedad del lado derecho de la ecuación de modelo (12.1) se encuentra en el error aleatorio ϵ y su valor medio y varianza son 0 y σ^2, respectivamente, cualquiera que sea el valor de x. Esto implica que

$$\mu_{Y\cdot x^*} = E(\beta_0 + \beta_1 x^* + \epsilon) = \beta_0 + \beta_1 x^* + E(\epsilon) = \beta_0 + \beta_1 x^*$$
$$\sigma^2_{Y\cdot x^*} = V(\beta_0 + \beta_1 x^* + \epsilon) = V(\beta_0 + \beta_1 x^*) + V(\epsilon) = 0 + \sigma^2 = \sigma^2$$

Reemplazando x^* en $\mu_{Y\cdot x^*}$ por x se obtiene la relación $\mu_{Y\cdot x} = \beta_0 + \beta_1 x$, la cual expresa que el *valor medio de Y*, en lugar de Y misma, es una función lineal de x. La línea de regresión verdadera $y = \beta_0 + \beta_1 x$ es por consiguiente la *línea de valores medios*; su altura por encima de cualquier valor x es el valor esperado de Y para ese valor de x. La pendiente β_1 de la línea de regresión verdadera se interpreta como el cambio *esperado* de Y asociado con el incremento en 1 unidad del valor de x. La segunda relación manifiesta que la cantidad de variabilidad en la distribución de valores Y es la misma con cada valor diferente de x (homogeneidad de varianza). En el ejemplo que implica la edad de un niño y el tamaño de su vocabulario, el modelo implica que el tamaño de vocabulario promedio cambia linealmente con la edad (afortunadamente β_1 es positiva) y que la cantidad de variabilidad del tamaño de vocabulario a cualquier edad particular es la misma que a cualquier otra edad. Por último, con x fija, Y es la suma de una constante $\beta_0 + \beta_1 x$ y una variable aleatoria normalmente distribuida ϵ así que tiene una distribución normal. Estas propiedades se ilustran en la figura 12.4. El parámetro de varianza σ^2 determina el grado al cual

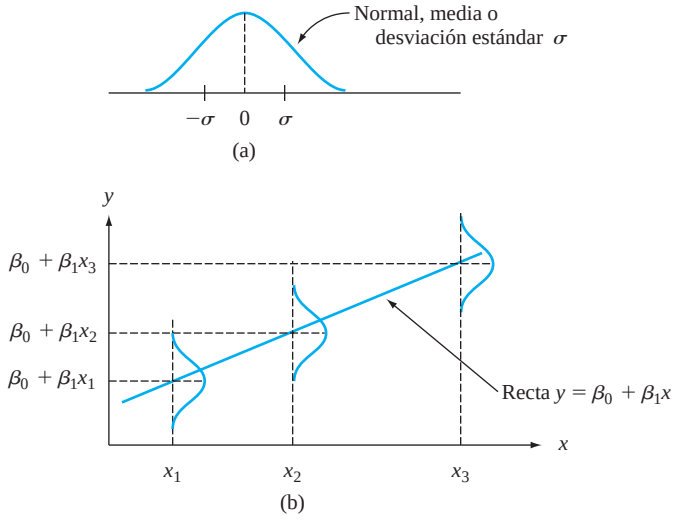


Figura 12.4 (a) Distribución de ϵ; (b) distribución de Y con diferentes valores de x

cada curva normal se dispersa en torno a su valor medio (la altura de la línea). Cuando σ^2 es pequeña, un punto observado (x, y) normalmente quedará bastante cerca de la línea de regresión verdadera mientras que las observaciones pueden desviarse considerablemente de sus valores esperados (correspondientes a puntos alejados de la línea) cuando σ^2 es grande.

Ejemplo 12.3 Supóngase que el modelo de regresión lineal simple con una línea de regresión verdadera $y = 65 - 1.2x$ y $\sigma = 8$ describe la relación entre el esfuerzo aplicado x y el tiempo para la fractura y. Entonces con cualquier valor fijo x^* de esfuerzo, el tiempo para la fractura tiene una distribución normal con valor medio de $65 - 1.2x^*$ y desviación estándar 8. Crudamente hablando, en la población compuesta de todos los puntos (x, y), la magnitud de una desviación típica con respecto a la línea de regresión verdadera es aproximadamente 8. Con $x = 20$, Y tiene un valor medio $\mu_{Y \cdot 20} = 65 - 1.2(20) = 41$, por lo tanto

$$P(Y > 50 \text{ cuando } x = 20) = P\left(Z > \frac{50 - 41}{8}\right) = 1 - \Phi(1.13) = .1292$$

La probabilidad de que el tiempo para la fractura exceda de 50 cuando se aplica un esfuerzo de 25 es debido a que $\mu_{Y \cdot 25} = 35$,

$$P(Y > 50 \text{ cuando } x = 25) = P\left(Z > \frac{50 - 35}{8}\right) = 1 - \Phi(1.88) = .0301$$

Estas probabilidades se ilustran como las áreas sombreadas en la figura 12.5.

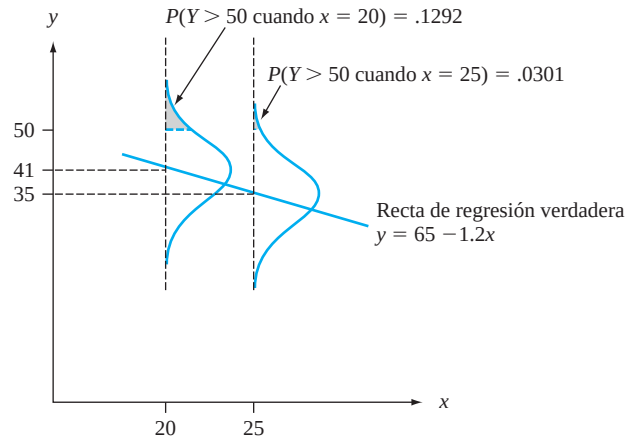


Figura 12.5 Probabilidades basadas en el modelo de regresión lineal simple

Supóngase que Y_1 denota una observación del tiempo para la fractura realizada con $x = 25$ y que Y_2 denota una observación independiente realizada con $x = 24$. Entonces $Y_1 - Y_2$ está normalmente distribuida con valor medio $E(Y_1 - Y_2) = \beta_1 = -1.2$, varianza $V(Y_1 - Y_2) = \sigma^2 + \sigma^2 = 128$ y desviación estándar $\sqrt{128} = 11.314$. La probabilidad de que Y_1 exceda a Y_2 es

$$P(Y_1 - Y_2 > 0) = P\left(Z > \frac{0 - (-1.2)}{11.314}\right) = P(Z > .11) = .4562$$

Es decir, aunque se espera que Y disminuya a medida que x se incrementa en 1 unidad, no es improbable que la Y observada en $x + 1$ será más grande que la Y observada en x. ■

EJERCICIOS Sección 12.1 (1–11)

1. La relación de eficiencia de un espécimen de acero sumergido en un tanque de fosfatado es el peso del recubrimiento de fosfato dividido entre la pérdida de metal (ambos en mg/pie^2). El artículo "Statistical Process Control of a Phosphate Coating Line" (*Wire J. Intl*., mayo de 1997: 78–81) dio los datos adjuntos sobre la temperatura del tanque (x) y relación de eficiencia (y):

Temp.	170	172	173	174	174	175	176
Relación	.84	1.31	1.42	1.03	1.07	1.08	1.04
Temp.	177	180	180	180	180	180	181
Relación	1.80	1.45	1.60	1.61	2.13	2.15	.84
Temp.	181	182	182	182	182	184	184
Relación	1.43	.90	1.81	1.94	2.68	1.49	2.52
Temp.	185	186	188				
Relación	3.00	1.87	3.08				

a. Construya gráficas de tallo y hojas tanto de la relación de temperatura como de la relación de eficiencia y comente características interesantes.

b. ¿Está el valor de la relación de eficiencia determinado por completo y de forma única por la temperatura del tanque? Explique su razonamiento.

c. Construya una gráfica de dispersión de los datos. ¿Parece que la relación de eficiencia podría ser pronosticada muy bien por el valor de la temperatura? Explique su razonamiento.

2. El artículo "Exhaust Emissions from Four-Stroke Lawn Mower Engines" (*J. of the Air and Water Mgmnt. Assoc*., 1997: 945–952) reportó datos de un estudio en el cual se utilizó una mezcla de gasolinas básicas y una gasolina reformulada. Considere las siguientes observaciones sobre edad (años) y emisiones de NO_x (g/kWh):

Motor	1	2	3	4	5
Edad	0	0	2	11	7
Línea de base	1.72	4.38	4.06	1.26	5.31
Reformulada	1.88	5.93	5.54	2.67	6.53
Motor	6	7	8	9	10
Edad	16	9	0	12	4
Línea de base	.57	3.37	3.44	.74	1.24
Reformulada	.74	4.94	4.89	.69	1.42

Construya gráficas de dispersión de emisiones de NO_x contra edad. ¿Cuál parece ser la naturaleza de la relación entre estas dos variables? [*Nota*: los autores del artículo citado comentaron sobre la relación.]

3. A menudo surgen datos bivariantes cuando se utilizan dos técnicas diferentes de medir la misma cantidad. Como un ejemplo, las observaciones adjuntas de x = concentración de hidrógeno (ppm) por medio de un método de cromatografía de gases y = concentración mediante un nuevo método de sensor se leyeron en una gráfica que aparece en el artículo "A New Method to Measure the Diffusible Hydrogen Content in Steel Weldments Using a Polymer Electrolyte-Based Hydrogen Sensor" (*Welding Res*., julio de 1997: 251s–256s).

x	47	62	65	70	70	78	95	100	114	118
y	38	62	53	67	84	79	93	106	117	116
x	124	127	140	140	140	150	152	164	198	221
y	127	114	134	139	142	170	149	154	200	215

Construya una gráfica de dispersión. ¿Parece haber una fuerte relación entre los dos tipos de mediciones de concentración? ¿Parece que los dos métodos miden aproximadamente la misma cantidad? Explique su razonamiento.

4. Un estudio para valorar la capacidad de sistemas de humedecimiento de suelos mediante flujo subsuperficial para eliminar la demanda de oxígeno bioquímico (BOD, por sus siglas en inglés) y varios otros constituyentes químicos dio los datos adjuntos sobre x = carga masiva de BOD (kg/ha/d) y y = eliminación masiva de BOD (kg/ha/d) ("Subsurface Flow Wetlands–A Performance Evaluation", *Water Envir. Res*., 1995: 244–247).

x	3	8	10	11	13	16	27	30	35	37	38	44	103	142
y	4	7	8	8	10	11	16	26	21	9	31	30	75	90

a. Construya gráficas de dispersión tanto de carga masiva como de eliminación masiva y comente sobre cualquier característica interesante.

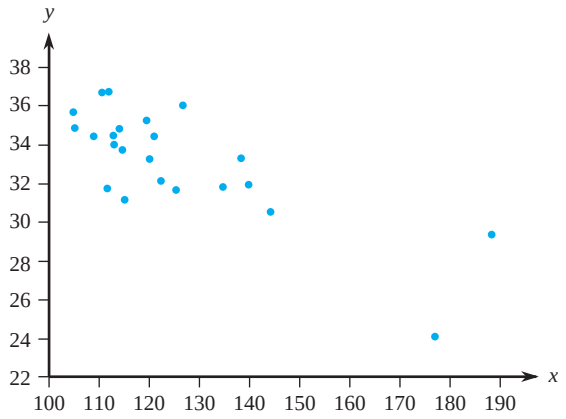
b. Construya una gráfica de dispersión de los datos y comente sobre cualquier característica importante.

5. El artículo "Objective Measurement of the Stretchability of Mozzarella Cheese" (*J. of Texture Studies*, 1992: 185–194) reportó sobre un experimento para investigar la variación del comportamiento del queso mozzarella con la temperatura. Considere los datos adjuntos sobre x = temperatura y y = alargamiento (%) en el momento de la falla del queso. [*Nota*: los investigadores eran italianos y utilizaron queso mozzarella *real*, no el pobre primo ampliamente disponible en Estados Unidos.]

x	59	63	68	72	74	78	83
y	118	182	247	208	197	135	132

a. Construya una gráfica de dispersión en la cual los ejes se corten en (0, 0). Marque 0, 20, 40, 60, 80 y 100 en el eje horizontal y 0, 50, 100, 150, 200 y 250 en el eje vertical.
b. Construya una gráfica de dispersión en la cual los ejes se corten en (55, 100), como se hizo en el citado artículo. ¿Parece ser preferible esta gráfica a la del inciso (a)? Explique su razonamiento.
c. ¿Qué sugieren las gráficas de los incisos (a) y (b) sobre la naturaleza de la relación entre las dos variables?

6. Un factor en el desarrollo del codo de tenista, una dolencia que provoca terror en el corazón de todos los tenistas serios, es la vibración inducida por el impacto del sistema raqueta y brazo al contacto con la pelota. Es bien sabido que la probabilidad de sufrir de codo de tenista depende de varias propiedades de la raqueta utilizada. Considere la gráfica de dispersión de x = frecuencia de resonancia de la raqueta y y = suma de la aceleración pico a pico (una característica de la vibración del brazo, en m/s/s) de $n = 23$ raquetas diferentes ("Transfer of Tennis Racket Vibrations into the Human Forearm", *Medicine and Science in Sports and Exercise*, 1992: 1134–1140). Discuta características interesantes de los datos y la gráfica de dispersión.



7. El artículo "Some Field Experience in the Use of an Accelerated Method in Estimating 28-Day Strength of Concrete" (*J. of Amer. Concrete Institute*, 1969: 895) consideró regresar y = resistencia estándar después de 28 días de curado (lb/pulg^2) contra x = resistencia acelerada (lb/pulg^2). Suponga que la ecuación de la línea de regresión verdadera es $y = 1800 + 1.3x$.

a. ¿Cuál es el valor esperado de la resistencia después de 28 días cuando la resistencia acelerada = 2500?
b. ¿En cuánto se debe esperar que cambie la resistencia después de 28 días cuando la resistencia acelerada se incrementa en 1 lb/pulg^2?
c. Responda la parte (b) para un incremento de 100 lb/pulg^2.
d. Responda la parte (b) para una reducción de 100 lb/pulg^2.

8. Recurriendo al ejercicio 7, suponga que la desviación estándar de la desviación aleatoria ϵ es de 350 lb/pulg^2.
a. ¿Cuál es la probabilidad de que el valor observado de la resistencia después de 28 días excederá de 5000 lb/pulg^2 cuando el valor de la resistencia acelerada es de 2000?
b. Repita la parte (a) con 2500 en lugar de 2000.
c. Considere hacer dos observaciones independientes de resistencia después de 28 días, la primera con una resistencia acelerada de 2000 y la segunda con $x = 2500$. ¿Cuál es la probabilidad de que la segunda observación excederá la primera por más de 1000 lb/pulg^2?
d. Sean Y_1 y Y_2 las observaciones de resistencia después de 28 días cuando $x = x_1$ y $x = x_2$, respectivamente. ¿Por cuánto tendría que exceder x_2 a x_1 para que $P(Y_2 > Y_1) = .95$?

9. La velocidad de flujo y (m^3/min) en un dispositivo utilizado para medir la calidad del aire depende de la caída de presión x (pulg. de agua) a través del filtro del dispositivo. Suponga que con valores de x entre 5 y 20, las dos variables están relacionadas de acuerdo con el modelo de regresión lineal simple con línea de regresión verdadera $y = -.12 + .095x$.
a. ¿Cuál es el cambio esperado de la velocidad de flujo asociado con un incremento de 1 pulg en la caída de presión? Explique.
b. ¿Qué cambio de la velocidad de flujo puede ser esperado cuando la caída de presión se reduce en 5 pulg?
c. ¿Cuál es la velocidad de flujo esperada con una caída de presión de 10 pulg? ¿Una caída de presión de 15 pulg?
d. Suponga $\sigma = .025$ y considere una caída de presión de 10 pulg. ¿Cuál es la probabilidad de que el valor observado de la velocidad de flujo excederá de .835?, ¿de que la velocidad de flujo observada excederá de .840?
e. ¿Cuál es la probabilidad de que una observación de la velocidad de flujo cuando la caída de presión es de 10 pulg excederá una observación de la velocidad de flujo cuando la caída de presión es de 11 pulg?

10. Suponga que el costo esperado de una corrida de producción está relacionado con el tamaño de la corrida por conducto de la ecuación $y = 4000 + 10x$. Sea Y una observación sobre el costo de una corrida. Si el *tamaño* de las variables y el *costo* están relacionados de acuerdo con el modelo de regresión lineal simple, ¿podría ser el caso que $P(Y > 5500$ cuando $x = 100) = .05$ y $P(Y > 6500$ cuando $x = 200) = .10$? Explique.

11. Suponga que en un cierto proceso químico el tiempo de reacción y (h) está relacionado con la temperatura (°F) en la cámara en la cual la reacción ocurre de acuerdo con el modelo de regresión lineal simple con la ecuación $y = 5.00 - .01x$ y $\sigma = .075$.
a. ¿Cuál es el cambio esperado del tiempo de reacción con un incremento de 1°F de la temperatura? ¿Con un incremento de 10°F de la temperatura?

b. ¿Cuál es el tiempo de reacción esperado cuando la temperatura es de 200°F? ¿Cuando la temperatura es de 250°F?

c. Suponga que se realizan cinco observaciones independientemente del tiempo de reacción, cada una para una temperatura de 250°F. ¿Cuál es la probabilidad de que las cinco observaciones resulten entre 2.4 y 2.6 h?

d. ¿Cuál es la probabilidad de que dos tiempos de reacción independientemente observados a temperaturas con 1° de diferencia son tales que el tiempo a la temperatura más alta excede el tiempo a la temperatura más baja?

12.2 Estimación de parámetros de modelo

Se supondrá en ésta y en las siguientes secciones que las variables x y y están relacionadas de acuerdo con el modelo de regresión lineal simple. Un investigador casi nunca conocerá los valores de β_0, β_1 y σ^2. En cambio, estará disponible una muestra de datos compuesta de n pares observados $(x_1, y_1), \ldots, (x_n, y_n)$, con la cual los parámetros de modelo y la recta de regresión verdadera pueden ser estimados. Se supone que estas observaciones se obtuvieron independientemente una de otra. Es decir, y_1 es el valor observado de Y_i, donde $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ y las n desviaciones, $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ son variables aleatorias independientes. La independencia de $Y_1, Y_2, \ldots, Y_n$ se desprende de la independencia de las ϵ_i.

De acuerdo con el modelo, los puntos observados estarán distribuidos en torno a la recta de regresión verdadera de una manera aleatoria. La figura 12.6 muestra una gráfica típica de pares observados junto con dos candidatos para la recta de regresión estimada. Intuitivamente, la recta $y = a_0 + a_1 x$ no es una estimación razonable de la recta verdadera $y = \beta_0 + \beta_1 x$ porque, si $y = a_0 + a_1 x$ fuera la recta verdadera, los puntos observados con toda seguridad habrían quedado más cerca de esta línea. La recta $y = b_0 + b_1 x$ es una estimación más plausible porque los puntos observados están dispersos en lugar de estar cerca de esta recta.

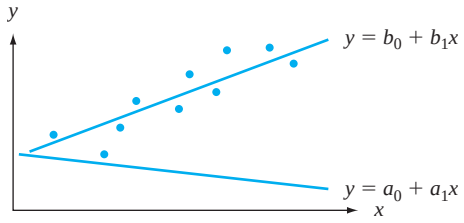


Figura 12.6 Dos estimaciones diferentes de la recta de regresión verdadera

La figura 12.6 y la discusión anterior sugieren que la estimación de $y = \beta_0 + \beta_1 x$ deberá ser una recta que en un cierto sentido se ajuste mejor a los puntos de los datos observados. Esto es lo que motiva el principio de mínimos cuadrados, el que puede ser rastreado hacia atrás en el tiempo hasta el matemático alemán Gauss (1777–1855). De acuerdo con este principio, una recta proporciona un buen ajuste para los datos si las distancias verticales (desviaciones) de los puntos observados a la línea son pequeñas (véase la figura 12.7). La medida de la bondad del ajuste es la suma de cuadrados de estas desviaciones. La recta de mejor ajuste es entonces la que tiene la suma más pequeña posible de desviaciones al cuadrado.

.

Principio de mínimos cuadrados

La desviación vertical del punto (x_i, y_i) con respecto a la línea $y = b_0 + b_1x$ es

$$\text{la altura del punto} - \text{altura de la línea} = y_1 - (b_0 + b_1x_i)$$

La suma de las desviaciones verticales al cuadrado de los puntos $(x_1, y_1), \ldots, (x_n, y_n)$ a la línea es entonces

$$f(b_0, b_1) = \sum_{i=1}^{n} [y_i - (b_0 + b_1x_i)]^2$$

Las estimaciones puntuales de β_0 y β_1, denotadas por $\hat{\beta}_0$ y $\hat{\beta}_1$ llamadas **estimaciones de mínimos cuadrados,** son aquellos valores que reducen al mínimo a $f(b_0, b_1)$. Es decir, $\hat{\beta}_0$ y $\hat{\beta}_1$ son tales que $f(\hat{\beta}_0, \hat{\beta}_1) \leq f(b_0, b_1)$ con cualesquier b_0 y b_1. La **recta de regresión estimada** o **recta de mínimos cuadrados** es entonces la recta cuya ecuación es $y = \hat{\beta}_0 + \hat{\beta}_1x$.

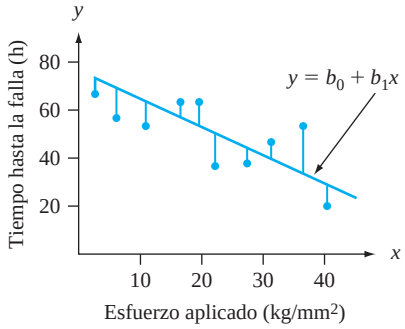


Figura 12.7 Desviaciones de los datos observados con respecto a la recta $y = b_0 + b_1x$

Los valores minimizados de b_0 y b_1 se encuentran tomando las derivadas parciales de $f(b_0, b_1)$ con respecto tanto a b_0 como b_1, igualándolas a cero [análogamente a $f'(b) = 0$ en cálculo univariante] y resolviendo las ecuaciones

$$\frac{\partial f(b_0, b_1)}{\partial b_0} = \sum 2(y_i - b_0 - b_1x_i)(-1) = 0$$
$$\frac{\partial f(b_0, b_1)}{\partial b_1} = \sum 2(y_i - b_0 - b_1x_i)(-x_i) = 0$$

Al cancelar el factor -2 y reordenar se obtiene el siguiente sistema de ecuaciones, llamado **ecuaciones normales:**

$$nb_0 + (\sum x_i)b_1 = \sum y_i$$
$$(\sum x_i)b_0 + (\sum x_i^2)b_1 = \sum x_iy_i$$

Estas ecuaciones son lineales en las dos incógnitas b_0 y b_1. Siempre que no todas las x_i sean idénticas, las estimaciones de mínimos cuadrados son la única solución de este sistema.

La estimación de mínimos cuadrados del coeficiente de pendiente β_1 de la recta de regresión verdadera es

$$b_1 = \hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}} \qquad (12.2)$$

Calculando fórmulas para el numerador y denominador de $\hat{\beta}_1$ son

$$S_{xy} = \sum x_i y_i - (\sum x_i)(\sum y_i)/n \qquad S_{xx} = \sum x_i^2 - (\sum x_i)^2/n$$

La estimación de mínimos cuadrados de la intersección β_0 de la línea de regresión verdadera es

$$b_0 = \hat{\beta}_0 = \frac{\sum y_i - \hat{\beta}_1 \sum x_i}{n} = \bar{y} - \hat{\beta}_1 \bar{x} \qquad (12.3)$$

Las fórmulas para el cálculo de S_{xy} y S_{xx} requieren sólo los estadísticos resumidos $\sum x_i$, $\sum y_i$, $\sum x_i^2$ y $\sum x_i y_i$ ($\sum y_i^2$ se requerirá en breve). Al calcular $\hat{\beta}_0$ se utilizan dígitos adicionales en $\hat{\beta}_1$ porque, si $\bar{x}$ es grande en magnitud, el redondeo afectará la respuesta final. En la práctica, el uso de un paquete de software estadístico es preferible al cálculo manual y gráficos dibujados a mano. Una vez más, asegúrese de que el gráfico de dispersión muestre un patrón lineal con una variación relativamente homogénea antes de ajustar el modelo de regresión lineal simple.

Ejemplo 12.4 El número de cetano es una propiedad fundamental en la especificación de la calidad de ignición del combustible utilizado en un motor diesel. La determinación de este número para un combustible biodiesel es cara y lleva mucho tiempo. El artículo "Relating the Cetane Number of Biodiesel Fuels to Their Fatty Acid Composition: A Critical Study" (*J. of Automobile Engr.*, 2009: 565–583) incluye los siguientes datos en x = índice de yodo (g) y y = número de cetano para una muestra de 14 biocombustibles. El índice de yodo es la cantidad de yodo necesario para saturar una muestra de 100 g de aceite. Los autores del artículo ajustan el modelo de regresión lineal simple a estos datos, así que vamos a seguir su ejemplo.

x	132.0	129.0	120.0	113.2	105.0	92.0	84.0	83.2	88.4	59.0	80.0	81.5	71.0	69.2
y	46.0	48.0	51.0	52.1	54.0	52.0	59.0	58.7	61.6	64.0	61.4	54.6	58.8	58.0

El resumen de las cantidades necesarias para el cálculo manual se puede obtener mediante la colocación de los valores de x en una columna y los valores y en otra columna y, a continuación la creación de columnas para x^2, xy y y^2 (estos últimos valores no son necesarios en el momento pero se utilizarán en breve). Calculando las sumas por columna, tenemos $\sum x_i = 1307.5$, $\sum y_i = 779.2$, $\sum x_i^2 = 128{,}913.93$, $\sum x_i y_i = 71{,}347.30$, $\sum y_i^2 = 43{,}745.22$, de donde

$$S_{xx} = 128{,}913.93 - (1307.5)^2/14 = 6802.7693$$

$$S_{xy} = 71{,}347.30 - (1307.5)(779.2)/14 = -1424.41429$$

La pendiente estimada de la recta de regresión real (es decir, la pendiente de la recta de mínimos cuadrados) es

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{-1424.41429}{6802.7693} = -.20938742$$

Estimamos que el cambio esperado en el promedio real del número de cetano asociado con un incremento de 1 g en el índice de yodo es −.209, es decir, una disminución de .209. Como $\bar{x} = 93.392857$ y $\bar{y} = 55.657143$, es la intersección estimada de la recta de regresión real (es decir, la intersección de la recta de mínimos cuadrados) es

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 55.657143 - (-.20938742)(93.392857) = 75.212432$$

La ecuación de la recta de regresión estimada (recta de mínimos cuadrados) es, $y = 75.212 - .2094x$ exactamente la descrita en el artículo citado. La figura 12.8 muestra un diagrama de dispersión de los datos con la recta de mínimos cuadrados sobrepuesta. Esta recta ofrece un resumen muy bueno de la relación entre las dos variables.

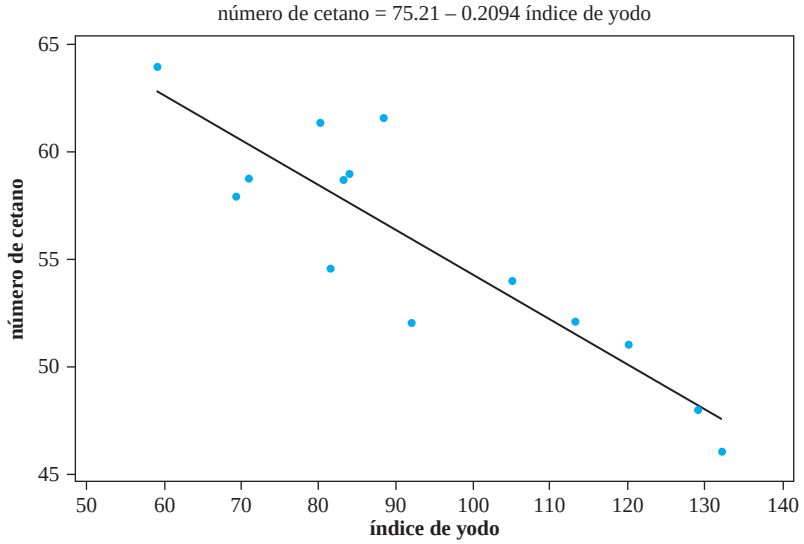


Figura 12.8 Diagrama de dispersión de los datos con la recta de mínimos cuadrados sobrepuesta con Minitab para el ejemplo 12.4

La recta de regresión estimada puede ser utilizada de inmediato para dos propósitos diferentes. Con un valor fijo de x, x^*, $\hat{\beta}_0 + \hat{\beta}_1 x^*$ (la altura de la recta sobre x^*) da (1) una estimación puntual del valor esperado de Y cuando $x = x^*$ o (2) una predicción puntual del valor Y que resultará de una nueva observación realizada con $x = x^*$.

Ejemplo 12.5 Remítase al escenario del valor del índice de yodo para el número de cetano descrito en el ejemplo anterior. La ecuación de regresión estimada fue $y = 75.212 - .2094x$. Una estimación puntual del verdadero número de cetano promedio de todos los biocombustibles, cuyo índice de yodo es 100 es

$$\hat{\mu}_{Y \cdot 100} = \hat{\beta}_0 + \hat{\beta}_1(100) = 75.212 - .2094(100) = 54.27$$

Si se selecciona una muestra de biocombustible cuyo índice de yodo es 100, también 54.27 es un punto de predicción para el número de cetano resultante

La recta de mínimos cuadrados no deberá ser utilizada para predecir un valor de x mucho más allá del rango de los datos, de tal suerte que $x = 40$ o $x = 150$ en el ejemplo 12.4. El **peligro de extrapolación** es que la relación ajustada (una recta en este caso) puede no ser válida para tales valores de x.

Estimación de σ^2 y σ

El parámetro σ^2 determina la cantidad de variabilidad, inherente en el modelo de regresión. Un valor grande de σ^2 conducirá a (x_i, y_i) observados que están bastante dispersos en torno a la línea de regresión verdadera, mientras que σ^2 sea pequeña los puntos observados tenderán a quedar cerca de la recta verdadera (véase la figura 12.9). Se utilizará una estimación de σ^2 en fórmulas de intervalos de confianza (IC) y procedimientos de prueba de hipótesis presentados en las dos secciones siguientes. Como la ecuación de la recta verdadera es desconocida, la estimación se basa en el grado al cual las observaciones muestrales se desvían de la recta estimada. Muchas desviaciones grandes (residuos) sugieren un valor grande de σ^2, mientras que las desviaciones de pequeña magnitud sugieren que σ^2 es pequeña.

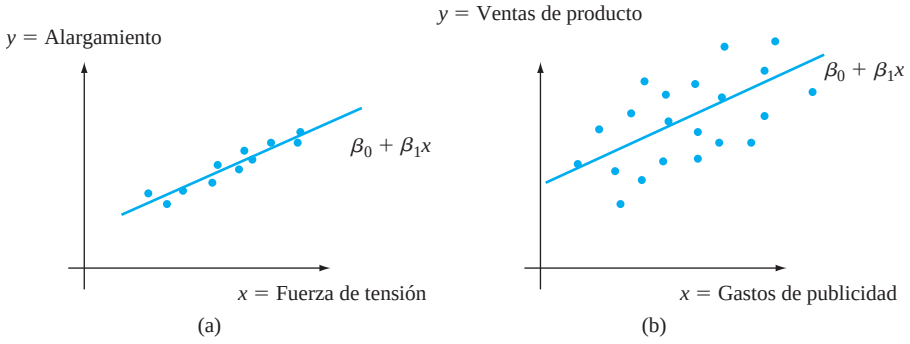


Figura 12.9 Muestra típica para σ^2: (a) pequeña; (b) grande

DEFINICIÓN

Los **valores ajustados** (o **pronosticados**) $\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n$ se obtienen sustituyendo sucesivamente $x_1, \ldots, x_n$ en la ecuación de la recta de regresión estimada: $\hat{y}_1 = \hat{\beta}_0 + \hat{\beta}_1 x_1, \hat{y}_2 = \hat{\beta}_0 + \hat{\beta}_1 x_2, \ldots, \hat{y}_n = \hat{\beta}_0 + \hat{\beta}_1 x_n$. Los **residuos** son las diferencias $y_1 - \hat{y}_1, y_2 - \hat{y}_2, \ldots, y_n - \hat{y}_n$ entre los valores observados y los valores ajustados.

En palabras, el valor pronosticado $\hat{y}_i$ es el valor de y pronosticado o esperado cuando se utiliza la recta de regresión estimada con $x = x_i$; $\hat{y}_i$ es la altura de la línea de regresión estimada por encima del valor x_i con el cual se realizó la i-ésima observación. El residuo $y_i - \hat{y}_i$ es la desviación vertical entre el punto (x_i, y_i) y la recta de mínimos cuadrados, un número positivo si el punto está sobre la recta y negativo si está debajo de ésta. Si todos los residuos son pequeños, entonces mucha de la variabilidad en los valores y observados parece deberse a la relación lineal entre x y y, mientras que muchos residuos grandes sugieren un poco de variabilidad inherente en y con respecto a la cantidad debida a la relación lineal. Suponiendo que la recta en la figura 12.7 es la recta de mínimos cuadrados, los residuos están identificados por segmentos de línea verticales que parten de los puntos observados a la recta. Cuando se obtiene la recta de regresión estimada vía el principio de mínimos cuadrados, la suma de los residuos en teoría debe ser cero. En la práctica, la suma puede desviarse un poco de cero debido al redondeo.

Ejemplo 12.6 La alta densidad de población de Japón ha provocado un sinnúmero de problemas de consumo de recursos. Una dificultad especialmente seria tiene que ver con la eliminación de desechos. El artículo "Innovative Sludge Handling Through Pelletization Thickening" (*Water Research*, 1999: 3245–3252) reportó el desarrollo de una nueva máquina de compresión para procesar lodos de albañal. Una parte importante de la investigación implicó

relacionar el contenido de humedad de gránulos comprimidos (y, en %) con la velocidad de filtración de la máquina (x, en kg-DS/m/h). Los siguientes datos se tomaron de una gráfica incluida en el artículo.

x	125.3	98.2	201.4	147.3	145.9	124.7	112.2	120.2	161.2	178.9
y	77.9	76.8	81.5	79.8	78.2	78.3	77.5	77.0	80.1	80.2
x	159.5	145.8	75.1	151.4	144.2	125.0	198.8	132.5	159.6	110.7
y	79.9	79.0	76.7	78.2	79.5	78.1	81.5	77.0	79.0	78.6

Las cantidades resumidas pertinentes (*estadísticos resumidos*) son $\sum x_i = 2817.9$, $\sum y_i = 1574.8$, $\sum x_i^2 = 415{,}949.85$, $\sum x_i y_i = 222{,}657.88$ y $\sum y_i^2 = 124{,}039.58$ de donde $\bar{x} = 140.895$, $\bar{y} = 78.74$, $S_{xx} = 18{,}921.8295$ y $S_{xy} = 776.434$. Por lo tanto

$$\hat{\beta}_1 = \frac{776.434}{18{,}921.8295} = .04103377 \approx .041$$

$$\hat{\beta}_0 = 78.74 - (.04103377)(140.895) = 72.958547 \approx 72.96$$

por lo que la ecuación de la recta de mínimos cuadrados es $y = 72.96 + .041x$. Para precisión numérica, los valores ajustados se calcularon con $\hat{y}_i = 72.958547 + .04103377x_i$:

$$\hat{y}_1 = 72.958547 + .04103377(125.3) \approx 78.100,\ y_1 - \hat{y}_1 \approx -.200, \text{ etc.}$$

Nueve de los 20 residuos son negativos, por lo que los nueve puntos correspondientes en un diagrama de dispersión de los datos, se encuentran por debajo de la recta de regresión estimada. Todos los valores previstos (ajustes) y residuos aparecen en la tabla adjunta.

Obs	Filtrado	Contenido de humedad	Ajuste	Residuo
1	125.3	77.9	78.100	−0.200
2	98.2	76.8	76.988	−0.188
3	201.4	81.5	81.223	0.277
4	147.3	79.8	79.003	0.797
5	145.9	78.2	78.945	−0.745
6	124.7	78.3	78.075	0.225
7	112.2	77.5	77.563	−0.063
8	120.2	77.0	77.891	−0.891
9	161.2	80.1	79.573	0.527
10	178.9	80.2	80.299	−0.099
11	159.5	79.9	79.503	0.397
12	145.8	79.0	78.941	0.059
13	75.1	76.7	76.040	0.660
14	151.4	78.2	79.171	−0.971
15	144.2	79.5	78.876	0.624
16	125.0	78.1	78.088	0.012
17	198.8	81.5	81.116	0.384
18	132.5	77.0	78.396	−1.396
19	159.6	79.0	79.508	−0.508
20	110.7	78.6	77.501	1.099

Casi de la misma forma en que las desviaciones de la media en una situación de una muestra se combinaron para obtener la estimación $s^2 = \sum(x_i - \bar{x})^2/(n-1)$, la estimación de σ^2 en un análisis de regresión se basa en elevar al cuadrado y sumar los residuos. Se continuará utilizando el símbolo s^2 para esta varianza estimada, así que no hay que confundirla con la s^2 previa.

DEFINICIÓN

La **suma de cuadrados debido al error** (o de forma equivalente, suma de cuadrados residuales) denotada por SSE, es

$$\text{SSE} = \sum(y_i - \hat{y}_i)^2 = \sum[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2$$

y la estimación de σ^2 es

$$\hat{\sigma}^2 = s^2 = \frac{\text{SSE}}{n-2} = \frac{\sum(y_i - \hat{y}_i)^2}{n-2}$$

El divisor $n - 2$ en s^2 es el número de grados de libertad (gl) asociado con la estimación (o, de forma equivalente), con la suma de cuadrados debido al error y la estimación s^2. Esto es porque para obtener s^2, primero se deben estimar los dos parámetros β_0 y β_1, lo que hace que se pierdan 2 grados de libertad (exactamente como μ hubo de ser estimada en problemas de una muestra, con el resultado de una varianza estimada basada en $n - 1$ grados de libertad). Sustituyendo cada y_i en la fórmula para s^2 por la variable aleatoria Y_i, tenemos el estimador S^2. Puede demostrarse que S^2 es un estimador insesgado para σ^2 (aunque el estimador S no sea insesgado para σ. Una interpretación de s que es similar a lo que se sugirió anteriormente para la desviación estándar muestral: A grandes rasgos, es el tamaño de una desviación típica vertical dentro de la muestra desde la recta de regresión estimada.

Ejemplo 12.7 Previamente se calcularon los residuos de los datos de contenido de humedad–velocidad de filtración. La suma de cuadrados debido al error correspondiente es

$$\text{SSE} = (-.200)^2 + (-.188)^2 + \cdots + (1.099)^2 = 7.968$$

La estimación de σ^2 es entonces $\hat{\sigma}^2 = s^2 = 7.968/(20-2) = .4427$ y la desviación estándar estimada es $\hat{\sigma} = s = \sqrt{.4427} = .665$. Crudamente hablando, .665 es la magnitud de una desviación típica con respecto a la recta de regresión estimada, algunos puntos están cerca de la recta tanto como otros están alejados. ■

El cálculo de SSE con la fórmula definitoria implica mucha aritmética tediosa porque primero se deben calcular los valores pronosticados y residuos. La siguiente fórmula de cálculo no requiere estas cantidades.

$$\text{SSE} = \sum y_i^2 - \hat{\beta}_0 \sum y_i - \hat{\beta}_1 \sum x_i y_i$$

Esta expresión se obtiene al sustituir $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ en $\sum(y_i - \hat{y}_i)^2$, elevar al cuadrado el sumando, realizar la suma y continuarla hasta los tres términos resultantes y simplificar. La fórmula de cálculo es especialmente sensible a los efectos de redondeo en $\hat{\beta}_0$ y $\hat{\beta}_1$, así que conservar tantos dígitos como sea posible en cálculos intermedios protegerá contra errores de redondeo.

Ejemplo 12.8 El artículo "Promising Quantitative Nondestructive Evaluation Techniques for Composite Materials" (*Materials Evaluation*, 1985: 561–565) reporta sobre un estudio para investigar cómo la propagación de una onda de esfuerzo ultrasónica a través de una sustancia depende de las propiedades de ésta. Los datos adjuntos sobre resistencia a la fractura (x, como porcentaje de resistencia a la tensión última) y atenuación (y, en neper/cm, la disminución de la amplitud de la onda de esfuerzo) en compuestos de poliéster reforzados con fibra de vidrio se tomaron de una gráfica que aparece en el artículo. El patrón lineal sustancial que aparece en la gráfica de dispersión sugiere el modelo de regresión lineal simple.

x	12	30	36	40	45	57	62	67	71	78	93	94	100	105
y	3.3	3.2	3.4	3.0	2.8	2.9	2.7	2.6	2.5	2.6	2.2	2.0	2.3	2.1

Las cantidades resumidas necesarias son $n = 14$, $\sum x_i = 890$, $\sum x_i^2 = 67{,}182$, $\sum y_i = 37.6$, $\sum y_i^2 = 103.54$ y $\sum x_i y_i = 2234.30$ de donde $S_{xx} = 10{,}603.4285714$, $S_{xy} = -155.98571429$, $\hat{\beta}_1 = -.0147109$ y $\hat{\beta}_0 = 3.6209072$. Entonces

$$\begin{aligned} \text{SSE} &= 103.54 - (3.6209072)(37.6) - (-.0147109)(2234.30) \\ &= .2624532 \end{aligned}$$

por lo tanto $s^2 = .2624532/12 = .0218711$ y $s = .1479$. Cuando $\hat{\beta}_0$ y $\hat{\beta}_1$ se redondean a tres cifras decimales en la fórmula de cálculo para SSE, el resultado es

$$\text{SSE} = 103.54 - (3.621)(37.6) - (-.015)(2234.30) = .905$$

la cual es más de tres veces el valor correcto. ■

Coeficiente de determinación

La figura 12.10 muestra tres gráficas de dispersión diferentes de datos bivariantes. En las tres gráficas, las alturas de los diferentes puntos varían sustancialmente, lo que indica que existe mucha variabilidad en los valores y observados. Todos los puntos en la primera gráfica quedan exactamente en una línea recta. En este caso toda la variación (el 100%) de y puede ser atribuida al hecho de que x y y están linealmente relacionadas en combinación con la variación de x. Los puntos en la figura 12.10(b) no quedan exactamente en una recta, pero su variabilidad se compara a la variabilidad total de y, las desviaciones con respecto a la recta de mínimos cuadrados son pequeñas. Es razonable concluir en este caso que una gran parte de la variación de y observada puede ser atribuida a la relación lineal aproximada entre las variables postuladas por el modelo de regresión lineal simple. Cuando la gráfica de dispersión es como la de la figura 12.10(c), existe una variación sustancial en torno a la recta de mínimos cuadrados con respecto a la variación total de y, así que el modelo de regresión lineal simple no explica la variación de y relacionando y con x.

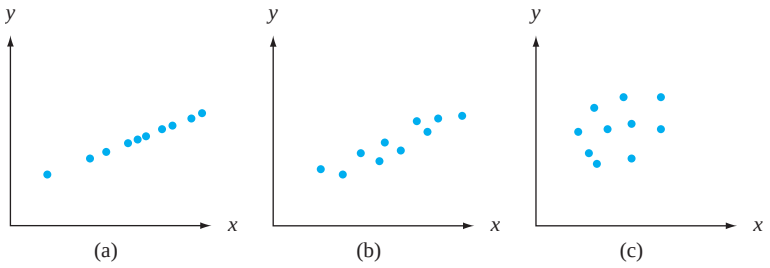


Figura 12.10 Utilización del modelo para explicar la variación de y: (a) datos para los que se explica toda la variación; (b) datos con los cuales la mayor parte de la variación es explicada; (c) datos con los cuales poca variación es explicada.

La suma de cuadrados SSE debido al error puede ser interpretada como una medida de cuánta variación de y permanece sin ser explicada por el modelo; es decir, cuánta no puede ser atribuida a una relación lineal. En la figura 12.10(a), SSE $= 0$ y no existe ninguna variación no explicada, en tanto ésta sea pequeña con los datos de la figura 12.10(b) y mucho más grande en la figura 12.10(c). La **suma de cuadrados total** da una medida cuantitativa de la cantidad de variación total en los valores y observados

$$\text{SST} = S_{yy} = \Sigma(y_i - \bar{y})^2 = \Sigma y_i^2 - (\Sigma y_i)^2/n$$

La suma de cuadrados total es la suma de las desviaciones al cuadrado con respecto a la media muestral de los valores y observados. Por consiguiente se resta el mismo número $\bar{y}$ de cada y_i presente en SST, mientras que SSE implica restar cada valor diferente pronosticado $\hat{y}_i$ de la y_i correspondiente observada. Así como SSE es la suma de desviaciones al cuadrado con respecto a la recta de mínimos cuadrados $y = \hat{\beta}_0 + \hat{\beta}_1 x$, SST es la suma de desviaciones al cuadrado con respecto a la línea horizontal a la altura $\bar{y}$ (en tal caso las desviaciones verticales son $y_i - \bar{y}$), como se ilustra en la figura 12.11. Además, como la suma de desviaciones al cuadrado con respecto a la recta de mínimos cuadrados es más pequeña que la suma de desviaciones al cuadrado con respecto a *cualquier* otra línea, SSE $<$ SST a menos que la línea horizontal misma sea la recta de mínimos cuadrados. La razón SSE/SST es la proporción de variación total que no puede ser explicada por el modelo de regresión lineal simple y 1 $-$ SSE/SST (un número entre 0 y 1) es la proporción de variación de y observada explicada por el modelo.

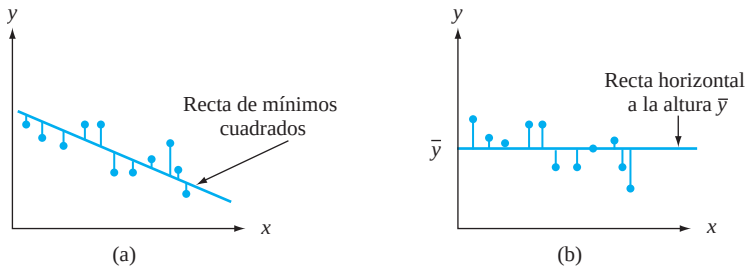


Figura 12.11 Sumas de cuadrados ilustrada: (a) SSE = suma de desviaciones cuadráticas en torno a la recta de mínimos cuadrados; (b) SST = suma de desviaciones al cuadrado en torno a la línea horizontal

DEFINICIÓN

El **coeficiente de determinación,** denotado por r^2, está dado por

$$r^2 = 1 - \frac{\text{SSE}}{\text{SST}}$$

Se interpreta como la proporción de variación y observada que puede ser explicada por el modelo de regresión lineal simple (atribuida a una relación lineal aproximada entre y y x).

Mientras más alto es el valor de r^2, más exitoso es el modelo de regresión lineal simple al explicar la variación de y. Cuando se realiza un análisis de regresión mediante un programa estadístico, r^2 o $100r^2$ (el porcentaje de variación explicado por el modelo) es una parte prominente de los resultados. Si r^2 es pequeño, un analista normalmente deseará buscar un modelo alternativo (ya sea un modelo no lineal o uno de regresión múltiple que implique más de una sola variable independiente) que explique con más eficacia la variación de y.

Ejemplo 12.9 La gráfica de dispersión de los datos para el valor del índice de yodo del número de cetano de la figura 12.8 ciertamente pretende un valor de r^2 razonablemente alto. Con

$$\hat{\beta}_0 = 75.212432 \quad \hat{\beta}_1 = -.20938742 \quad \sum y_i = 779.2$$
$$\sum x_i y_i = 71{,}347.30 \quad \sum y_i^2 = 43{,}745.22$$

se tiene

$$\text{SST} = 43{,}745.22 - (779.2)^2/14 = 377.174$$

$$\text{SSE} = 43{,}745.22 - (75.212432)(779.2) - (-.20938742)(71{,}347.30) = 78.920$$

El coeficiente de determinación es entonces

$$r^2 = 1 - \text{SSE/SST} = 1 - (78.920)/(377.174) = .791$$

Esto es, el 79.1% de la variación observada en el número de cetano es atribuible a (puede ser explicada por) la regresión lineal simple que relaciona el número de cetano y el valor del índice de yodo (los valores de r^2 son incluso más altos que esto en muchos contextos científicos, pero los científicos sociales generalmente estarían extasiados con ¡un valor grande cerca de éste!)

La figura 12.12 muestra resultados parciales generados por Minitab con los datos del número de cetano en el índice de yodo. El programa también proporciona los valores y residuos pronosticados si se los solicita, así como también otra información. Los formatos utilizados por otros programas difieren un poco de los de Minitab, pero el contenido de la información es muy similar. La suma de cuadrados de regresión se estudiará en breve. Las demás cantidades en la figura 12.12 que aún no han sido discutidas saldrán a la superficie en la sección 12.3 [excepto R-Sq(adj), que entra al juego en el capítulo 13 cuando se estudian los modelos de regresión múltiple].

```
La ecuación de regresión es
núm cetano  = 75.2 − 0.209 índice yodo

                                β̂0   β̂1
Predictor          Coef      SE Coef          T         P
Constante        75.212        2.984      25.21     0.000
índice yodo    −0.20939      0.03109     −6.73     0.000
                                      100r²
s = 2.56450  R-sq = 79.1%   R-sq(adj) = 77.3%

Análisis de varianza              SSE

FUENTE         DF        SS          MS          F         P
Regresión       1    298.25      298.25      45.35     0.000
Error          12     78.92        6.58
Total          13    377.17 ←————————————  SST
```

Figura 12.12 Resultados obtenidos con Minitab para la regresión de los ejemplos 12.4 y 12.9. ■

El coeficiente de determinación se escribe en una forma un poco diferente al introducir una tercera suma de cuadrados; la **suma de cuadrados debida a la regresión.** SSR-dada por $\text{SSR} = \sum(\hat{y}_i - \bar{y})^2 = \text{SST} - \text{SSE}$. La suma de cuadrados debido a la regresión se interpreta como la cantidad de variación total que *es* explicada por el modelo. En tal caso se tiene

$$r^2 = 1 - \text{SSE/SST} = (\text{SST} - \text{SSE})/\text{SST} = \text{SSR/SST}$$

la relación de la variación explicada a la variación total. La tabla ANOVA que aparece en la figura 12.12 muestra que SSR = 298.25 de donde $r^2 = 298.25/377.17 = .791$ como antes.

Terminología y alcance del análisis de regresión

El término *análisis de regresión* fue introducido por primera vez por Francis Galton a finales del siglo XIX en conexión con su trabajo sobre la relación entre la estatura del padre x y la estatura del hijo y. Después de recopilar varios pares (x_i, y_i), Galton utilizó el principio de mínimos cuadrados para obtener la ecuación de la línea de regresión estimada con el objetivo de utilizarla para predecir la estatura del hijo a partir de la estatura del padre. Al utilizar la recta obtenida, Galton encontró que si la estatura del padre era por encima del promedio, era de esperarse que la estatura del hijo también fuera por encima del promedio, *pero no tanto como el padre.* Asimismo, el hijo de un padre con una estatura más corta que el promedio también tendría una estatura más corta que el promedio, pero no tanto como el padre. Por lo tanto la estatura pronosticada de un hijo era "llevada de vuelta" hacia la media; porque *regresión* significa un regreso o venida. Galton adoptó la terminología *recta de regresión*. Este fenómeno de ser llevado de vuelta hacia la media ha sido observado en muchas otras situaciones (p. ej., promedios de bateo de un año al otro en el beisbol) y se llama **efecto de regresión.**

La discusión hasta ahora ha supuesto que la variable independiente está bajo el control del investigador, así que sólo la variable dependiente Y es aleatoria. Éste no fue, sin embargo, el caso con el experimento de Galton: las estaturas de los padres no fueron preseleccionadas, sino que en su lugar tanto X como Y fueron aleatorias. Se pueden aplicar métodos y conclusiones de análisis de regresión cuando los valores de la variable independiente se fijan de antemano y cuando son aleatorios, pero como las derivaciones e interpretaciones son más directas en el primer caso, se continuará trabajando explícitamente con él. Para más comentarios, véase el excelente libro de John Neter y colaboradores, citado en la bibliografía del capítulo.

EJERCICIOS Sección 12.2 (12–29)

12. El ejercicio 4 dio datos sobre carga masiva de x = BOD y eliminación masiva de y = BOD. Valores de cantidades resumidas pertinentes son

$n = 14 \quad \sum x_i = 517$

$\sum y_i = 346 \quad \sum x_i^2 = 39{,}095$

$\sum y_i^2 = 17{,}454 \quad \sum x_i y_i = 25{,}825$

a. Obtenga una ecuación de la recta de mínimos cuadrados.
b. Pronostique el valor de la eliminación masiva de BOD de una sola observación realizada cuando la carga masiva de BOD es de 35, y calcule el valor del residuo correspondiente.
c. Calcule SSE y luego una estimación puntual de σ.
d. ¿Qué proporción de la variación observada de la eliminación puede ser explicada por la relación lineal aproximada entre las dos variables?
e. Los últimos dos valores de x, 103 y 142, son mucho más grandes que los demás. ¿Cómo se ven afectados la ecuación de la recta de mínimos cuadrados y el valor de r^2 por la supresión de las dos observaciones correspondientes de la muestra? Ajuste los valores dados de las cantidades resumidas y use el hecho de que el nuevo valor de SSE es 311.79.

13. Los datos adjuntos sobre x = densidad de corriente (mA/cm^2) y y = tasa de deposición (μm/min) aparecieron en el artículo "Plating of 60/40 Tin/Lead Solder for Head Termination Metallurgy" (*Plating and Surface Finishing*, enero de 1997: 38–40). ¿Está de acuerdo en que la afirmación del autor del artículo de que "se obtuvo una relación lineal a partir de la tasa de deposición de estaño-plomo como una función de la densidad de corriente?" Explique su razonamiento.

x	20	40	60	80
y	.24	1.20	1.71	2.22

14. Remítase a los datos de relación de temperatura del tanque-eficiencia dada en el ejercicio 1.
a. Determine la ecuación de la línea de regresión estimada.
b. Calcule una estimación puntual de la relación eficiencia promedio verdadera cuando la temperatura del tanque es de 182.
c. Calcule los valores de los residuos con la recta de mínimos cuadrados de las cuatro observaciones con las cuales la temperatura es de 182. ¿Por qué no todas tienen el mismo signo?
d. ¿Qué proporción de la variación observada en la relación de eficiencia puede ser atribuida a la relación de regresión lineal simple entre las dos variables?

15. Se determinaron valores de módulo de elasticidad (MOE, la relación de esfuerzo, es decir, fuerza por unidad de área a deformación por unidad de longitud, en GPa) y resistencia a la flexión (una medida de la capacidad para resistir la falla en la flexión en MPa) con una muestra de vigas de concreto de un cierto tipo, y se obtuvieron los siguientes datos (tomados de una gráfica que aparece en el artículo "Effects of Aggregates and Microfillers on the Flexural Properties of Concrete", *Magazine of Concrete Research*, 1997: 81–98):

MOE	29.8	33.2	33.7	35.3	35.5	36.1	36.2
Resistencia	5.9	7.2	7.3	6.3	8.1	6.8	7.0
MOE	36.3	37.5	37.7	38.7	38.8	39.6	41.0
Resistencia	7.6	6.8	6.5	7.0	6.3	7.9	9.0
MOE	42.8	42.8	43.5	45.6	46.0	46.9	48.0
Resistencia	8.2	8.7	7.8	9.7	7.4	7.7	9.7
MOE	49.3	51.7	62.6	69.8	79.5	80.0	
Resistencia	7.8	7.7	11.6	11.3	11.8	10.7	

a. Construya una gráfica de tallo y hojas de los valores de MOE y comente sobre cualquier característica interesante.

b. ¿Es el valor de resistencia completa y únicamente determinado por el valor del MOE? Explique.

c. Use los resultados adjuntos generados por Minitab para obtener la ecuación de la recta de mínimos cuadrados para predecir resistencia a partir del módulo de elasticidad y luego para predecir resistencia de una viga cuyo módulo de elasticidad es de 40. ¿Se sentiría cómodo si utiliza la recta de mínimos cuadrados para predecir resistencia cuando el módulo de elasticidad es de 100? Explique.

```
Predictor   Coef     Stdev  t-ratio      P
Constante 3.2925   0.6008     5.48  0.000
mod elas 0.10748  0.01280     8.40  0.000

s = 0.8657  R-sq = 73.8%  R-sq(adj) = 72.8%

Análisis de varianza

FUENTE    DF      SS      MS      F       P
Regresión  1  52.870  52.870  70.55  0.000
Error     25  18.736   0.749
Total     26  71.605
```

d. ¿Cuáles son los valores de SSE, SST y el coeficiente de determinación? ¿Sugieren estos valores que el modelo de regresión lineal simple describe de forma efectiva la relación entre las dos variables? Explique.

16. El artículo "Characterization of Highway Runoff in Austin, Texas, Area" (*J. of Envir. Engr.*, 1998: 131–137) incluye una gráfica de dispersión junto con una recta de mínimos cuadrados, de x = volumen de precipitación pluvial (m^3) y y = volumen de escurrimiento (m^3) en un lugar particular. Los valores adjuntos se tomaron de la gráfica.

x	5	12	14	17	23	30	40	47
y	4	10	13	15	15	25	27	46

x	55	67	72	81	96	112	127
y	38	46	53	70	82	99	100

a. ¿Apoya una gráfica de dispersión de los datos el uso del modelo de regresión lineal simple?

b. Calcule estimaciones puntuales de la pendiente e intersección de la recta de regresión de población.

c. Calcule una estimación puntual del volumen de escurrimiento promedio verdadero cuando el volumen de precipitación pluvial es de 50.

d. Calcule una estimación puntual de la desviación estándar σ.

e. ¿Qué proporción de la variación observada del volumen de escurrimiento puede ser atribuida a la relación de regresión lineal simple entre el escurrimiento y la precipitación pluvial?

17. El agregado fino de concreto, hecho a partir de un agregado secundario clasificado de manera uniforme y una pasta de cemento-agua, es beneficioso en las zonas propensas a las lluvias excesivas debido a sus excelente propiedades de drenaje. El artículo "Pavement Thickness Design for No-Fines Concrete Parking Lots", *J. of Trans. Engr.*, 1995: 476–484) empleó un análisis de mínimos cuadrados en el estudio de cómo y = porosidad (%) se relaciona con x = peso unitario (por pie cúbico) en muestras de concreto. Considere los datos siguientes representativos:

x	99.0	101.1	102.7	103.0	105.4	107.0	108.7	110.8
y	28.8	27.9	27.0	25.2	22.8	21.5	20.9	19.6

x	112.1	112.4	113.6	113.8	115.1	115.4	120.0
y	17.1	18.9	16.0	16.7	13.0	13.6	10.8

Un resumen de las cantidades pertinentes es $\sum x_i = 1640.1$, $\sum y_i = 299.8$, $\sum x_i^2 = 179{,}849.73$, $\sum x_i y_i = 32{,}308.59$, $\sum y_i^2 = 6430.06$.

a. Obtenga la ecuación de la recta de regresión estimada. A continuación, cree un gráfico de dispersión de los datos y el gráfico de la recta estimada. ¿Parece que el modelo de la relación puede explicar una gran parte de la variación observada en y?

b. Interprete la pendiente de la recta de mínimos cuadrados.

c. ¿Qué sucede si la estimación lineal se utiliza para predecir la porosidad cuando el peso unitario es de 135? ¿Por qué no es una buena idea?

d. Calcule los residuos correspondientes a las dos primeras observaciones.

e. Calcule e interprete una estimación puntual de σ.

f. ¿Qué proporción de la variación observada en la porosidad se puede atribuir a la relación lineal aproximada entre el peso unitario y la porosidad?

18. Durante la última década, el polvo de caucho se ha utilizado en cemento asfáltico para mejorar el rendimiento. El artículo "Experimental Study of Recycled Rubber-Filled High-Strength Concrete" (*Magazine of Concrete Res.*, 2009: 549–556) incluye una regresión de y = *esfuerzo axial* (MPa) en x = *esfuerzo cúbico* (Mpa) basada en los siguientes datos de muestra:

x	112.3	97.0	92.7	86.0	102.0	99.2	95.8	103.5	89.0	86.7
y	75.0	71.0	57.7	48.7	74.3	73.3	68.0	59.3	57.8	48.5

a. Obtenga la ecuación de la recta de mínimos cuadrados, e interprete su pendiente.

b. Calcule e interprete el coeficiente de determinación.

c. Calcule e interprete una estimación de la desviación estándar σ del error en el modelo de regresión lineal simple.

19. Los siguientes datos son representativos de los reportados en el artículo "An Experimental Correlation of Oxides of Nitrogen Emissions from Power Boilers Based on Field Data" (*J. of Engr. for Power*, julio de 1973: 165–170), con x = tasa de liberación debido a área de quemador (MBtu/h-pie^2) y y = tasa de emisión de NO_x (ppm):

x	100	125	125	150	150	200	200
y	150	140	180	210	190	320	280
x	250	250	300	300	350	400	400
y	400	430	440	390	600	610	670

a. Suponiendo que el modelo de regresión lineal simple es válido, obtenga la estimación de mínimos cuadrados de la línea de regresión verdadera.

b. ¿Cuál es la estimación de la tasa de emisión de NO_x esperada cuando la tasa de liberación debido al área del quemador es igual a 225?

c Estime la cantidad en la cual espera que cambie la tasa de emisiones de NO_x cuando la tasa de liberación debido al área del quemador disminuye en 50.

d. ¿Utilizaría la línea de regresión estimada para predecir la tasa de emisión con una tasa de liberación de 500? ¿Por qué sí o por qué no?

20. Varios estudios han demostrado que los líquenes (ciertas plantas compuestas de un alga y un hongo) son excelentes bioindicadores de la contaminación del aire. El artículo "The Epiphytic Lichen Hypogymnia Physodes as a Biomonitor of Atmospheric Nitrogen and Sulphur Deposition in Norway" (*Envir. Monitoring and Assesment*, 1993: 27–47) da los siguientes datos (tomados de una gráfica) sobre x = deposición de NO_3^- en húmedo (g N/m^2) y y = liquen (% de peso en seco):

x	.05	.10	.11	.12	.31	.37	.42
y	.48	.55	.48	.50	.58	.52	1.02
x	.58	.68	.68	.73	.85	.92	
y	.86	.86	1.00	.88	1.04	1.70	

El autor utilizó regresión lineal simple para analizar los datos. Use los resultados obtenidos con Minitab para responder las siguientes preguntas:

a. ¿Cuáles son las estimaciones de mínimos cuadrados de β_0 y β_1?

b. Pronostique el N de liquen con un valor de deposición de NO_3^- de .5.

c. ¿Cuál es la estimación de σ?

d. ¿Cuál es el valor de la variación total y cuánta de ella puede ser explicada por la relación de modelo?

```
La ecuación de regresión es
liquen N = 0.365 + 0.967 no3 depo

Predictor      Coef     Stdev   t-ratio       P
Constante   0.36510   0.09904      3.69   0.004
no3 depo     0.9668    0.1829      5.29   0.000

s = 0.1932    R-sq = 71.7%    R-sq(adj) = 69.2%

Análisis de varianza
FUENTE       DF       SS       MS       F       P
Regresión     1   1.0427   1.0427   27.94   0.000
Error        11   0.4106   0.0373
Total        12   1.4533
```

21. El ángulo de recuperación de arrugas y la resistencia a la tensión son las dos características más importantes para evaluar el desempeño de tela de algodón entrelazada. Un incremento en el ángulo de entrelazado, determinado por la absorbancia de una banda de éster carboxilo, mejora la resistencia a las arrugas de la tela (a expensas de reducir la resistencia mecánica). Los datos adjuntos sobre x = absorbancia y y = resistencia al ángulo de arrugamiento se tomaron de una gráfica incluida en el artículo "Predicting the Performance of Durable Press Finished Cotton Fabric with Infrared Spectroscopy" (*Textile Res. J.*, 1999: 145–151).

x	.115	.126	.183	.246	.282	.344	.355	.452	.491	.554	.651
y	334	342	355	363	365	372	381	392	400	412	420

He aquí los resultados obtenidos con Minitab:

```
Predictor       Coef     SE Coef         T         P
Constante    321.878       2.483    129.64     0.000
absorb       156.711       6.464     24.24     0.000

S = 3.60498   R-Sq = 98.5%   R-Sq(adj) = 98.3%

Fuente           DF       SS       MS       F       P
Resistencia       1   7639.0   7639.0  587.81   0.000
Error residual    9    117.0     13.0
Total            10   7756.0
```

a. ¿Parece ser apropiado el modelo de regresión lineal simple? Explique.

b. ¿Qué ángulo de resistencia a las arrugas pronosticaría para un espécimen de tela cuya absorbancia es de .300?

c. ¿Cuál sería la estimación del ángulo de resistencia a las arrugas esperado cuando la absorbancia es de .300?

22. El cemento de fosfato de calcio está ganando cada vez más atención para su uso en aplicaciones de reparación ósea. El artículo "Short-Fibre Reinforcement of Calcium Phosphate Bone Cement" (*J. of Engr. in Med.*, 2007: 203–211) informó sobre un estudio en el que se utilizan las fibras de polipropileno en un intento de mejorar el comportamiento de la fractura. Los siguientes datos de x = peso de la fibra (%) y y = resistencia a la compresión (MPa) fue proporcionado por los autores del artículo.

x	0.00	0.00	0.00	0.00	0.00	1.25	1.25	1.25	1.25
y	9.94	11.67	11.00	13.44	9.20	9.92	9.79	10.99	11.32
x	2.50	2.50	2.50	2.50	2.50	5.00	5.00	5.00	5.00
y	12.29	8.69	9.91	10.45	10.25	7.89	7.61	8.07	9.04
x	7.50	7.50	7.50	7.50	10.00	10.00	10.00	10.00	
y	6.63	6.43	7.03	7.63	7.35	6.94	7.02	7.67	

a. Ajuste el modelo de regresión lineal simple a estos datos. Después, determine la proporción de la variación observada en la resistencia que se puede atribuir a la relación entre el modelo de resistencia y el peso de la fibra. Por último, obtenga una estimación puntual de la desviación estándar de ϵ, la desviación aleatoria de la ecuación de modelo.

b. Los valores de resistencia promedio de los seis niveles diferentes de peso de la fibra son 11.05, 10.51, 10.32, 8.15, 6.93 y 7.24, respectivamente. El citado documento incluyó una figura en la que se comparó la resistencia promedio contra el peso promedio de la fibra. Obtenga la ecuación de esta recta de regresión y calcule el coeficiente de determinación

correspondiente. Explique la diferencia entre el valor de r^2 para esta regresión y el valor de r^2 obtenido en (a).

23. **a.** Obtenga la SSE con los datos del ejercicio 19 a partir de la fórmula definitoria [SSE = $\sum(y_i - \hat{y}_i)^2$] y compare con el valor determinado con la fórmula de cálculo.

b. Calcule el valor de la suma de cuadrados total. ¿Explica el modelo de regresión lineal la variación de la tasa de emisiones? Justifique su aseveración.

24. Los datos adjuntos se tomaron de una gráfica que apareció en el artículo "Reactions on Painted Steel Under the Influence of Sodium Chloride, and Combinations Thereof" (*Ind. Engr. Chem. Prod. Res. Dev.*, 1985: 375–378). La variable independiente es la tasa de deposición de SO_2 ($mg/m^2/d$) y la variable dependiente es pérdida de peso del acero (g/m^2).

x	14	18	40	43	45	112
y	280	350	470	500	560	1200

a. Construya una gráfica de dispersión. ¿Parece razonable el modelo de la regresión lineal simple en esta situación?

b. Calcule la ecuación de la recta de regresión estimada.

c. ¿Qué porcentaje de la variación observada en la pérdida de peso del acero puede ser atribuido a la relación de modelo en combinación con la variación de la tasa de deposición?

d. Debido a que el valor x más grande en la muestra excede en gran medida a los demás, esta observación puede haber influido mucho al determinar la ecuación de la recta estimada. Elimine esta observación y recalcule la ecuación. ¿Difiere la nueva ecuación sustancialmente de la original (podría considerar valores pronosticados)?

25. Compruebe que b_1 y b_0 de las expresiones (12.2) y 12.3) satisfacen las ecuaciones normales.

26. Demuestre que el "promedio de puntos" $(\bar{x}, \bar{y})$ queda en la recta de regresión estimada.

27. Suponga que un investigador cuenta con datos sobre la cantidad de espacio de anaquel x dedicado a la exhibición de un producto particular e ingresos por ventas y de ese producto. Puede que el investigador desee adaptar un modelo para el cual la recta de regresión verdadera pase a través de (0, 0). El modelo apropiado es $Y = \beta_1 x + \epsilon$. Suponga que $(x_1, y_1), \ldots, (x_n, y_n)$ son pares observados generados con este modelo y deduzca el estimador de mínimos cuadrados de β_1. [*Sugerencia*: escriba la suma de desviaciones al cuadrado como una función de b_1, un valor de prueba y use el cálculo para determinar el valor minimizante de b_1.]

28. **a.** Considere los datos del ejercicio 20. Suponga que en lugar de la recta de mínimos cuadrados que pasa por los puntos $(x_1, y_1), \ldots, (x_n, y_n)$, se desea que la recta de mínimos cuadrados pase por $(x_1 - \bar{x}, y_1), \ldots, (x_n - \bar{x}, y_n)$. Construya una gráfica de dispersión con los puntos (x_i, y_i) y luego con los puntos $(x_i - \bar{x}, y_i)$. Use las gráficas para explicar intuitivamente cómo están relacionadas entre sí las dos rectas de mínimos cuadrados.

b. Suponga que en lugar del modelo $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i (i = 1, \ldots, n)$, se desea ajustar un modelo de la forma $Y_i = \beta_0^* + \beta_1^*(x_i - \bar{x}) + \epsilon_i \left(i = 1, \ldots, n\right)$. ¿Cuáles son los estimadores de mínimos cuadrados de β_0^* y β_1^* y cómo están relacionados con $\hat{\beta}_0$ y $\hat{\beta}_1$?

29. Considere los siguientes tres conjuntos de datos, en los cuales las variables de interés son x = distancia de la casa al trabajo y y = tiempo para recorrer la distancia de la casa al trabajo. Basado en una gráfica de dispersión y los valores de s y r^2, ¿en qué situación la regresión lineal simple sería más (menos) efectiva y por qué?

Conjunto de datos	**1**		**2**		**3**	
	x	***y***	***x***	***y***	***x***	***y***
	15	42	5	16	5	8
	16	35	10	32	10	16
	17	45	15	44	15	22
	18	42	20	45	20	23
	19	49	25	63	25	31
	20	46	50	115	50	60
S_{xx}	17.50		1270.8333		1270.8333	
S_{xy}	29.50		2722.5		1431.6667	
$\hat{\beta}_1$	1.685714		2.142295		1.126557	
$\hat{\beta}_0$	13.666672		7.868852		3.196729	
SST	114.83		5897.5		1627.33	
SSE	65.10		65.10		14.48	

12.3 Inferencias sobre el parámetro de pendiente β_1

En virtualmente todo el trabajo inferencial realizado hasta ahora, la noción de variabilidad de muestreo ha sido persistente. En particular, las propiedades de las distribuciones de muestreo de varios estadísticos han sido la base para desarrollar fórmulas de intervalo de confianza y métodos de prueba de hipótesis. La idea clave en este caso es que el valor de cualquier cantidad calculada a partir de datos muestrales, el valor de cualquier estadístico, va a variar de una muestra a otra.

Ejemplo 12.10 Reconsidere los datos sobre x = tasa de liberación debido al área del quemador y y = tasa de emisiones de NO_x del ejercicio 12.19 en la sección previa. Existen 14 observaciones, realizadas con los valores x 100, 125, 125, 150, 150, 200, 200, 250, 250, 300, 300, 350, 400 y 400, respectivamente. Suponga que la pendiente e intersección de la línea de regresión verdadera son $\beta_1 = 1.70$ y $\beta_0 = -.50$ con $\sigma = 35$ (consistente con los valores $\hat{\beta}_1 = 1.7114$, $\hat{\beta}_0 = -45.55$, $s = 36.75$). Se procedió a generar una muestra de desviaciones aleatorias $\tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_{14}$ con respecto a una distribución normal con media 0 y desviación estándar 35 y luego se sumó $\tilde{\epsilon}_i$ a $\beta_0 + \beta_1 x_i$ para obtener 14 valores y correspondientes. Se realizaron entonces los cálculos de regresión para obtener la pendiente, la intersección y la desviación estándar estimados. Este proceso se repitió un total de 20 veces y los valores resultantes se dan en la tabla 12.1.

Tabla 12.1 Resultados de simulación del ejemplo 12.10

	$\hat{\beta}_1$	$\hat{\beta}_0$	s		$\hat{\beta}_1$	$\hat{\beta}_0$	s
1.	1.7559	−60.62	43.23	11.	1.7843	−67.36	41.80
2.	1.6400	−49.40	30.69	12.	1.5822	−28.64	32.46
3.	1.4699	−4.80	36.26	13.	1.8194	−83.99	40.80
4.	1.6944	−41.95	22.89	14.	1.6469	−32.03	28.11
5.	1.4497	5.80	36.84	15.	1.7712	−52.66	33.04
6.	1.7309	−70.01	39.56	16.	1.7004	−58.06	43.44
7.	1.8890	−95.01	42.37	17.	1.6103	−27.89	25.60
8.	1.6471	−40.30	43.71	18.	1.6396	−24.89	40.78
9.	1.7216	−42.68	23.68	19.	1.7857	−77.31	32.38
10.	1.7058	−63.31	31.58	20.	1.6342	−17.00	30.93

Claramente existe variación en los valores de la pendiente y la intersección estimadas, así como también la desviación estándar estimada. La ecuación de la recta de mínimos cuadrados varía por lo tanto de una muestra a la siguiente. La figura 12.13 en la página 492 muestra una gráfica de puntos de las pendientes estimadas así como también gráficas de la línea de regresión verdadera y las 20 líneas de regresión muestrales. ■

La pendiente β_1 de la recta de regresión de población es el cambio promedio verdadero en la variable dependiente y asociada con un incremento de 1 unidad en la variable independiente x. La pendiente de la recta de mínimos cuadrados, $\hat{\beta}_1$, da una estimación puntual de β_1. Del mismo modo que un intervalo de confianza para μ y los procedimientos para probar hipótesis con respecto a μ se basaron en propiedades de la distribución de muestreo de $\bar{X}$, las inferencias adicionales sobre β_1 están basadas en considerar a $\hat{\beta}_1$ como un estadístico e investigar su distribución de muestreo.

Se supone que los valores de las x_i se eligen antes de realizar el experimento, así que sólo las Y_i son aleatorias. Los estimadores (estadísticos, y por lo tanto variables aleatorias) de β_0 y β_1 se obtienen reemplazando y_i por Y_i en (12.2) y (12.3):

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(Y_i - \bar{Y})}{\sum(x_i - \bar{x})^2} \qquad \hat{\beta}_0 = \frac{\sum Y_i - \hat{\beta}_1 \sum x_i}{n}$$

Asimismo, el estimador de σ^2 se obtiene al reemplazar cada y_i en la fórmula para s^2 por la variable aleatoria Y_i:

$$\hat{\sigma}^2 = S^2 = \frac{\sum Y_i^2 - \hat{\beta}_0 \sum Y_i - \hat{\beta}_1 \sum x_i Y_i}{n - 2}$$

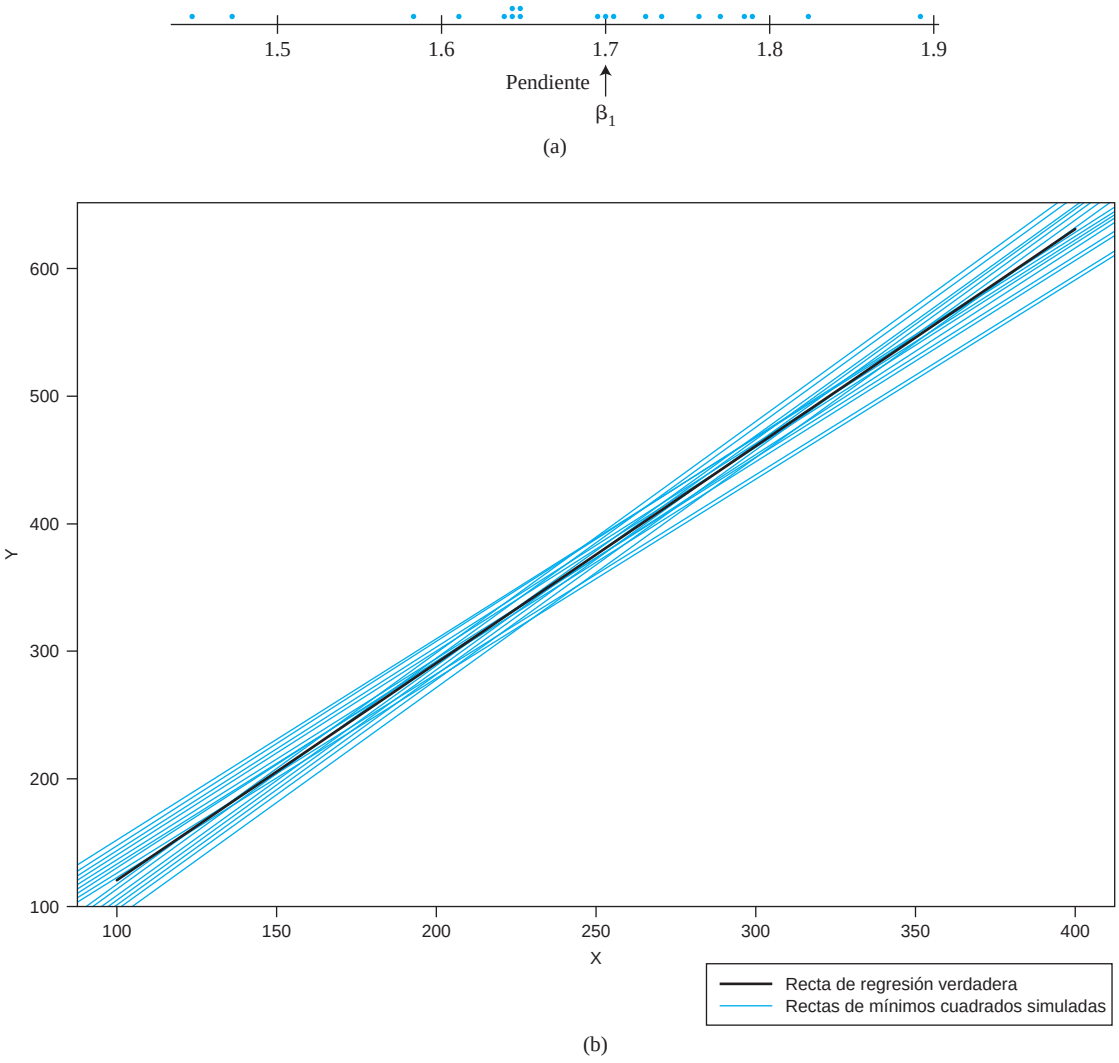


Figura 12.13 Resultados de simulación del ejemplo 12.10; (a) Gráfica de puntos de pendientes estimadas; (b) gráficas de la línea de regresión verdadera y 20 rectas de mínimos cuadrados (obtenidas con S-Plus)

El denominador de $\hat{\beta}_1$, $S_{xx} = \sum(x_i - \bar{x})^2$, depende sólo de las x_i y no de las Y_i, así que es una constante. Entonces como $\sum(x_i - \bar{x})\bar{Y} = \bar{Y}\sum(x_i - \bar{x}) = \bar{Y} \cdot 0 = 0$, el estimador de la pendiente se escribe como

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})Y_i}{S_{xx}} = \sum c_i Y_i \quad \text{donde } c_i = (x_i - \bar{x})/S_{xx}$$

Es decir, $\hat{\beta}_i$ es una función lineal de las variables aleatorias independientes $Y_1, Y_2, \ldots, Y_n$, cada una de las cuales está normalmente distribuida. Invocando las propiedades de una función lineal de variables aleatorias discutidas en la sección 5.5 conduce a los siguientes resultados.

PROPOSICIÓN

1. El valor medio de $\hat{\beta}_1$ es $E(\hat{\beta}_1) = \mu_{\hat{\beta}_1} = \beta_1$, así que $\hat{\beta}_1$ es un estimador insesgado de β_1 (la distribución de $\hat{\beta}_1$ siempre está centralizada en el valor de β_1).

2. La varianza y desviación estándar de $\hat{\beta}_1$ son

$$V(\hat{\beta}_1) = \sigma^2_{\hat{\beta}_1} = \frac{\sigma^2}{S_{xx}} \qquad \sigma_{\hat{\beta}_1} = \frac{\sigma}{\sqrt{S_{xx}}} \tag{12.4}$$

donde $S_{xx} = \sum(x_i - \bar{x})^2 = \sum x_i^2 - (\sum x_i)^2/n$. Reemplazando σ por su estimación s da una estimación para $\sigma_{\hat{\beta}_1}$ (la desviación estándar estimada, es decir, el error estándar estimado, de $\hat{\beta}_1$):

$$s_{\hat{\beta}_1} = \frac{s}{\sqrt{S_{xx}}}$$

(Esta estimación también puede ser denotada por $\hat{\sigma}_{\hat{\beta}_1}$.)

3. El estimador $\hat{\beta}_1$ tiene una distribución normal (porque es una función lineal de variables aleatorias estandarizadas independientes).

De acuerdo con (12.4), la varianza de $\hat{\beta}_1$ es igual a la varianza σ^2 del término de error aleatorio, o de forma equivalente, de cualquier Y_i,-dividida entre $\sum(x_i - \bar{x})^2$. Como $\bar{x}$ mide la dispersión de las x_i en torno a $\bar{x}$, se concluye que si se realizan observaciones a valores x_i que están bastante dispersos se obtiene un estimador más preciso del parámetro de pendiente (varianza más pequeña de $\hat{\beta}_1$), mientras que los valores de x_i muy cercanos entre sí implican un estimador altamente variable. Desde luego, si las x_i están demasiado dispersas, un modelo lineal puede no ser apropiado a lo largo del rango de observación.

Muchos procedimientos inferenciales previamente discutidos se basaron en estandarizar un estimador restando primero su valor medio y luego dividiéndolo entre su desviación estándar estimada. En particular, los procedimientos de prueba y un intervalo de confianza para μ media de una población normal utilizaron el hecho de que la variable estandarizada $(\overline{X} - \mu)/(S/\sqrt{n})$, es decir $(\overline{X} - \mu)/S_{\hat{\mu}}$, tenía una distribución t con $n-1$ grados de libertad. Un resultado similar en este caso abre la puerta a más inferencias sobre β_1.

TEOREMA

La suposición del modelo de regresión lineal simple implica que la variable estandarizada

$$T = \frac{\hat{\beta}_1 - \beta_1}{S/\sqrt{S_{xx}}} = \frac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}}$$

tiene una distribución t con $n - 2$ grados de libertad.

Un intervalo de confianza para β_1

Como en la deducción de intervalos de confianza previos, se inicia con un enunciado de probabilidad

$$P\left(-t_{\alpha/2,n-2} < \frac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}} < t_{\alpha/2,n-2}\right) = 1 - \alpha$$

La manipulación de las desigualdades entre los paréntesis para aislar β_1 y la sustitución de las estimaciones en lugar de los estimadores da la fórmula del intervalo de confianza.

Un **intervalo de confianza** de $100(1-\sigma)\%$ **para la pendiente** $\boldsymbol{\beta_1}$ de la recta de regresión verdadera es

$$\hat{\beta}_1 \pm t_{\alpha/2,n-2} \cdot s_{\hat{\beta}_1}$$

Este intervalo tiene la misma forma general de muchos de los intervalos previos. Está centrado en la estimación puntual del parámetro y la cantidad que se extiende a cada lado depende del nivel de confianza deseado (a través del valor crítico t) y de la cantidad de variabilidad del estimador $\hat{\beta}_1$ (a través de $s_{\hat{\beta}_1}$, el cual tenderá a ser más pequeño cuando existe poca variabilidad en la distribución de $\hat{\beta}_1$ y grande de lo contrario).

Ejemplo 12.11 Las variaciones del peso de mampostería de ladrillos de arcilla tienen implicaciones no sólo para diseño estructural y acústico sino también para el diseño de sistemas de calefacción, ventilación y aire acondicionado. El artículo "Clay Brick Masonry Weight Variation" (*J. of Architectural Engr.*, 1996: 135–137) incluye una gráfica de dispersión de y = densidad de mortero en seco (lb/pie^3) contra x = contenido de aire del mortero (%) para una muestra de especímenes de mortero, de donde se tomaron los siguientes datos representativos:

x	5.7	6.8	9.6	10.0	10.7	12.6	14.4	15.0	15.3
y	119.0	121.3	118.2	124.0	112.3	114.1	112.2	115.1	111.3
x	16.2	17.8	18.7	19.7	20.6	25.0			
y	107.2	108.9	107.8	111.0	106.2	105.0			

El diagrama de dispersión de estos datos en la figura 12.14 ciertamente sugiere la pertinencia del modelo de regresión lineal simple; parece haber una sustancial relación lineal negativa entre el contenido de aire y la densidad, una en la cual la densidad tiende a disminuir a medida que se incrementa el contenido de aire.

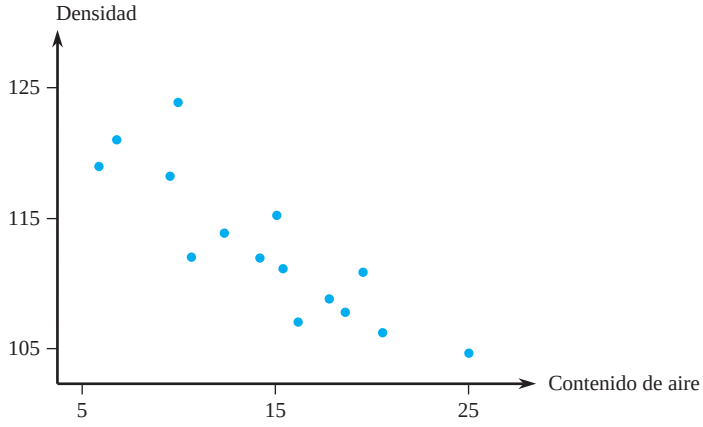


Figura 12.14 Diagrama de dispersión de los datos del ejemplo 12.11

Los valores de los estadísticos resumidos requeridos para calcular las estimaciones de mínimos cuadrados son

$$\sum x_i = 218.1 \quad \sum y_i = 1693.6 \quad \sum x_i^2 = 3577.01$$
$$\sum x_i y_i = 24{,}252.54 \qquad \sum y_i^2 = 191{,}672.90$$

con las cuales se obtuvieron $S_{xy} = -372.404$, $S_{xx} = 405.836$, $\hat{\beta}_1 = -.917622$, $\hat{\beta}_0 = 126.248889$, SST = 454.163, SSE = 112.4432 y $r^2 = 1 - 112.4432/454.1693 = .752$. Aproximadamente el 75% de la variación de la densidad observada puede ser atri-

buido a la relación de modelo de regresión lineal simple entre la densidad y el contenido de aire. Los grados de libertad debido al error es $15 - 2 = 13$, para obtener $s^2 = 112.4432/13 = 8.6495$ y $s = 2.941$.

La desviación estándar estimada de $\hat{\beta}_1$ es

$$s_{\hat{\beta}_1} = \frac{s}{\sqrt{S_{xx}}} = \frac{2.941}{\sqrt{405.836}} = .1460$$

Un nivel de confianza de 95% requiere $t_{.025,13} = 2.160$. El intervalo de confianza es

$$-.918 \pm (2.160)(.1460) = -.918 \pm .315 = (-1.233, -.603)$$

Con un alto grado de confianza, se estima que una disminución promedio de la densidad de entre .603 lb/pie^3 y 1.233 lb/pie^3 está asociada con un 1% de incremento del contenido de aire (por lo menos con valores de contenido de aire de entre aproximadamente 5% y 25%, correspondientes a los valores x de nuestra muestra). El intervalo es razonablemente angosto, lo que indica que la pendiente de la línea de población fue estimada con precisión. Obsérvese que el intervalo incluye sólo valores negativos, así que se puede estar seguro de la tendencia de la densidad a disminuir conforme el contenido de aire se incrementa.

Examinando los resultados obtenidos con SAS de la figura 12.15, se encuentra el valor de $s_{\hat{\beta}_1}$ bajo Estimaciones de parámetro como el segundo número en la columna Error estándar. Todos los programas estadísticos más ampliamente utilizados incluyen este error estándar estimado en el resultado. También hay un error estándar estimado para el estadístico $\hat{\beta}_0$ con el cual se puede calcular un intervalo de confianza para la intersección β_0 de la línea de regresión de la población.

Variable dependiente: DENSIDAD

Análisis de varianza

Fuente	GL	Suma de cuadrados	Media cuadrática	Valor F	Prob > F
Modelo	1	341.72606	341.72606	39.508	0.0001
Error	13	112.44327	8.64948		
C Total	14	454.16933			

Raíz MSE	2.94100	R-square	0.7524
Dep media	112.90667	Adj R-sq	0.7334
C.V.	2.60481		

Parámetros Estimados

Variable	DF	Parámetro estimado	Error estándar	T para H0: Parámetro=0	Prob >|T|
INTERSEC	1	126.248889	2.25441683	56.001	0.0001
CONT AIRE	1	−0.917622	0.14598888	−6.286	0.0001

Obs	Dep Var DENSIDAD	Valor predicho	Residual
1	119.0	121.0	−2.0184
2	121.3	120.0	1.2909
3	118.2	117.4	0.7603
4	124.0	117.1	6.9273
5	112.3	116.4	−4.1303
6	114.1	114.7	−0.5869
7	112.2	113.0	−0.8351
8	115.1	112.5	2.6154
9	111.3	112.2	−0.9093
10	107.2	111.4	−4.1834
11	108.9	109.9	−1.0152
12	107.8	109.1	−1.2894
13	111.0	108.2	2.8283
14	106.2	107.3	−1.1459
15	105.0	103.3	1.6917

Suma de residuos	0
Suma de residuales cuadráticos	112.4433
Residual predicho SS (presión)	146.4144

Figura 12.15 Resultados obtenidos con SAS con los datos del ejemplo 12.11 ■

Procedimientos de prueba de hipótesis

Como antes, la hipótesis nula en una prueba con respecto a β_1 será un enunciado de igualdad. El valor nulo (valor de β_1 supuesto verdadero por la hipótesis nula) será denotado por β_{10} (léase "beta uno cero", *no* "beta diez"). El estadístico de prueba se obtiene reemplazando β_1 en la variable estandarizada T por el valor nulo B_{10}; es decir, estandarizando el estimador de β_1 conforme a la suposición de que H_0 es verdadera. El estadístico de prueba tiene por lo tanto una distribución t con $n-2$ grados de libertad cuando H_0 es verdadera, así que la probabilidad de error de tipo I permanece al nivel deseado α utilizando un valor crítico t apropiado.

El par de hipótesis más comúnmente encontrado en torno a β_1 es H_0: $\beta_1 = 0$ contra H_a: $\beta_1 \neq 0$. Cuando esta hipótesis nula es verdadera, $\mu_{Y \cdot x} = \beta_0$ independiente de x, así que el conocimiento de x no da información sobre el valor de la variable dependiente. Una prueba de estas dos hipótesis a menudo se conoce como *prueba de utilidad del modelo* en regresión lineal simple. A menos que n sea demasiado pequeño, H_0 será rechazada y la utilidad del modelo confirmada con precisión cuando r^2 es razonablemente grande. El modelo de regresión lineal simple no deberá ser utilizado para más inferencias (estimaciones del valor medio o predicciones de valores futuros) a menos que la prueba de la utilidad del modelo dé por resultado el rechazo de H_0 con un α apropiadamente pequeño.

Hipótesis nula: H_0: $\beta_1 = \beta_{10}$

Valor estadístico de prueba: $t = \dfrac{\hat{\beta}_1 - \beta_{10}}{s_{\hat{\beta}_1}}$

Hipótesis alternativa	Región de rechazo para una prueba a nivel α
H_a: $\beta_1 > \beta_{10}$	$t \geq t_{\alpha,n-2}$
H_a: $\beta_1 < \beta_{10}$	$t \leq -t_{\alpha,n-2}$
H_a: $\beta_1 \neq \beta_{10}$	$t \geq t_{\alpha/2,n-2}$ o $t \leq -t_{\alpha/2,n-2}$

Se puede calcular un valor P basado en $n-2$ grados de libertad como previamente se hizo con pruebas t en los capítulos 8 y 9.

La **prueba de utilidad del modelo** es la prueba de H_0: $\beta_1 = 0$ contra H_a: $\beta_1 \neq 0$, en cuyo caso el valor estadístico de prueba es la **relación *t*** $t = \hat{\beta}_1 / s_{\hat{\beta}_1}$.

Ejemplo 12.12 Los ciclomotores o mopeds son muy populares en Europa debido a su costo y facilidad de operación. Sin embargo, pueden ser peligrosos si se modifican las características de rendimiento. Una de las características comúnmente manipulada es la velocidad máxima. El artículo "Procedure to Verify the Maximum Speed of Automatic Transmission Mopeds in Periodic Motor Vehicle Inspections" (*J. of Automotive Engr.*, 2008: 1615–1623) incluyó un análisis de regresión lineal simple de las variables x = velocidad de la pista de prueba (km/h) y y = velocidad de prueba de rodamiento. He aquí los datos leídos de un gráfico en el artículo:

x	42.2	42.6	43.3	43.5	43.7	44.1	44.9	45.3	45.7
y	44	44	44	45	45	46	46	46	47
x	45.7	45.9	46.0	46.2	46.2	46.8	46.8	47.1	47.2
y	48	48	48	47	48	48	49	49	49

Un gráfico de dispersión de los datos muestra un patrón lineal sustancial. La salida de Minitab en la figura 12.16 da el coeficiente de determinación como $r^2 = .923$, que sin duda presagia una relación lineal útil. Vamos a llevar a cabo la prueba de utilidad del modelo en un nivel de significancia $\alpha = .01$.

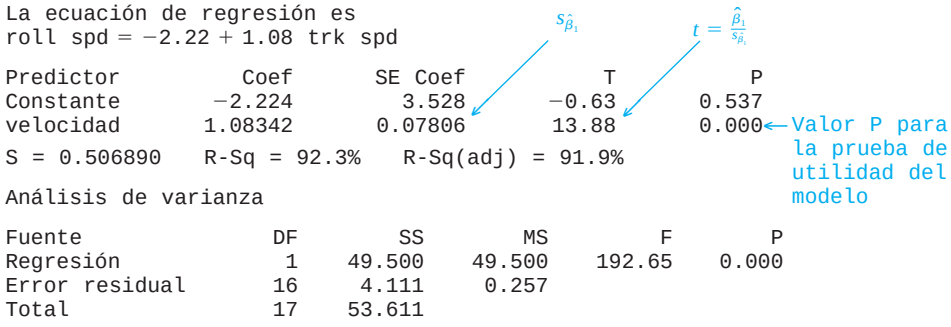


Figura 12.16 Resultados obtenidos con Minitab del ejemplo 12.12

El parámetro de interés es β_1, el cambio esperado en la velocidad de la prueba de rodado asociado con un incremento de 1 km/h en la velocidad de prueba. La hipótesis nula. H_0: $\beta_1 = 0$ será rechazada a favor de la alternativa H_0: $\beta_1 \neq 0$ si la relación t a $t = \hat{\beta}_1/s_{\hat{\beta}_1}$ satisface $t \geq t_{\alpha/2,n-2} = t_{.005,16} = 2.921$ o $t \leq 2.921$. De acuerdo con la figura 12.16, $\hat{\beta}_1 = 1.08342$, $s_{\hat{\beta}_1} = .07806$ y

$$t = \frac{1.08342}{.07806} = 13.88 \quad \text{(también en la salida de resultados)}$$

Claramente, esta razón t encaja bien en la cola superior de la región de rechazo de dos colas, de modo que H_0 es resonantemente rechazada. Alternativamente, el valor P es dos veces el área capturada bajo la curva t de 16 grados de libertad a la derecha de -11.11. Minitab da un valor $P = .000$, de modo que H_0 deberá ser rechazada a cualquier nivel α razonable. Esta confirmación de la utilidad de modelo de regresión simple permite calcular varias estimaciones y predicciones como se describe en la sección 12.4. ■

Regresión y ANOVA

La descomposición de la suma total de cuadrados $\sum(y_i - \bar{y})^2$ en una parte SSE, la cual mide la variación no explicada y una parte SSR, la cual mide la variación explicada por la relación lineal, hace recordar fuertemente el ANOVA unidireccional. De hecho, la hipótesis nula H_0: $\beta_1 = 0$ puede ser probada contra H_a: $\beta_1 \neq 0$ con una tabla ANOVA (tabla 12.2) y rechazando H_0 si $f \geq F_{\alpha,1,n-2}$.

La prueba F da exactamente el mismo resultado que la prueba t de utilidad de modelo $t^2 = f$ y $t^2_{\alpha/2,n-2} = F_{\alpha,1,n-2}$. Virtualmente todos los programas de computadora que cuentan con opciones de regresión incluyen tal tabla ANOVA en los resultados. Por ejemplo, la figura 12.15 muestra los resultados obtenidos con SAS con los datos de mortero del ejemplo 12.11. La tabla ANOVA en la parte superior de los resultados tiene $f = 39.508$ con un valor P de .0001 para la prueba de utilidad de modelo. La tabla de estimaciones de parámetro da $t = -6.286$ de nuevo con $P = .0001$ y $(-6.286)^2 = 39.51$.

Tabla 12.2 Tabla ANOVA para regresión lineal simple

Origen de la variación	Grados de libertad	Suma de cuadrados	Media cuadrática	f
Regresión	1	SSR	SSR	$\frac{\text{SSR}}{\text{SSE}/(n-2)}$
Error	$n - 2$	SSE	$s^2 = \frac{\text{SSE}}{n-2}$	
Total	$n - 1$	SST		

EJERCICIOS Sección 12.3 (30–43)

30. Reconsidere la situación descrita en el ejercicio 7, en el cual x = resistencia acelerada del concreto y y = resistencia después de 28 días de curado. Suponga que el modelo de regresión lineal simple es válido con x entre 1000 y 4000 y que $\beta_1 = 1.25$ y $\sigma = 350$. Considere un experimento en el cual $n = 7$ y los valores x a los cuales se realizan las observaciones son $x_1 = 1000$, $x_2 = 1500$, $x_3 = 2000$, $x_4 = 2500$, $x_5 = 3000$, $x_6 = 3500$, y $x_7 = 4000$.

- **a.** Calcule $\sigma_{\hat{\beta}_1}$, la desviación estándar de $\hat{\beta}_1$.
- **b.** ¿Cuál es la probabilidad de que la pendiente estimada basada en las observaciones será de entre 1.00 y 1.50?
- **c.** Suponga que también es posible hacer una sola observación con cada uno de los $n = 11$ valores $x_1 = 2000$, $x_2 = 2100$, . . . , $x_{11} = 3000$. Si un objetivo importante es estimar β_1 con tanta precisión como sea posible, ¿se preferiría el experimento con $n = 11$ a uno con $n = 7$?

31. Durante las operaciones de perforación de petróleo, los componentes del ensamble de perforación pueden sufrir de rompimiento por esfuerzo a partir de sulfuros. El artículo "Composition Optimization of High-Strength Steels for Sulfide Cracking Resistance Improvement" (*Corrosion Science*, 2009: 2878–2884) informó sobre un estudio en el que se analizó la composición de un acero de grado estándar. Los siguientes datos sobre el umbral de esfuerzo y = (SMYS%) y x = límite elástico (MPa) se leyeron de un gráfico en el artículo (que también incluye la ecuación de la recta de mínimos cuadrados)

x	635	644	711	708	836	820	810	870	856	923	878	937	948
y	100	93	88	84	77	75	74	63	57	55	47	43	38

$$\sum x_i = 10{,}576,\ \sum y_i = 894,\ \sum x_i^2 = 8{,}741{,}264,$$
$$\sum y_i^2 = 66{,}224,\ \sum x_i y_i = 703{,}192$$

- **a.** ¿Qué proporción de la variación observada en el esfuerzo puede ser atribuida a la relación lineal aproximada entre las dos variables?
- **b.** Calcule la desviación estándar estimada $s_{\hat{\beta}_1}$.
- **c.** Calcule un intervalo de confianza usando el nivel de confianza del 95% de la variación esperada del esfuerzo asociado con un aumento de 1 MPa en la resistencia. ¿Parece que este promedio real de cambio ha sido estimado con precisión?

32. El ejercicio 16 de la sección 12.2 dio datos sobre x = volumen de precipitación pluvial y y = volumen de escurrimiento (ambos en m^3). Use los resultados adjuntos obtenidos con Minitab para decidir si existe una relación lineal útil entre la precipitación pluvial y el escurrimiento y luego calcule un intervalo de confianza para el cambio promedio verdadero del volumen de escurrimiento asociado con 1 m^3 de incremento del volumen de precipitación pluvial.

```
La ecuación de regresión es
escurrimiento = −1.13 + 0.827 precipitación pluvial
Predictor          Coef      Stdev    t-ratio       P
Constante        −1.128      2.368     −0.48   0.642
precipitación   0.82697 0.03652 22.64   0.000
pluvial
s = 5.240     R-sq = 97.5%     R-sq(adj) = 97.3%
```

33. El ejercicio 15 de la sección 12.2 incluyó resultados generados por Minitab del módulo de elasticidad con una regresión de resistencia a la flexión de vigas de concreto.

- **a.** Úselos para calcular un intervalo de confianza con un nivel de confianza de 95% para la pendiente β_1 de la línea de regresión de población e interprete el intervalo resultante.
- **b.** Suponga que previamente se había creído que cuando el módulo de elasticidad se incrementa en 1 GPa, el cambio promedio verdadero asociado de la resistencia a la flexión era cuando mucho de .1 MPa. ¿Contradicen los datos esta creencia? Formule y pruebe las hipótesis pertinentes.

34. Remítase a los resultados generados por Minitab del ejercicio 20, en los cuales x = deposición de NO_3^- en húmedo y y = liquen N (%).

- **a.** Realice la prueba de utilidad de modelo al nivel .01, utilizando el método de región de rechazo.
- **b.** Repita el inciso (a) con el método del valor P.
- **c.** Suponga que previamente se creía que cuando la deposición en húmedo de NO_3^- se incrementa en .1 g N/m^2, el cambio asociado del liquen N esperado es por lo menos de .15%. Realice una prueba de hipótesis al nivel .01 para decidir si los datos contradicen esta creencia previa.

35. ¿Cómo afecta la aceleración lateral –fuerzas laterales experimentadas en las curvas que en gran medida están bajo el control del conductor– las náuseas percibidas por los pasajeros de un autobús? El artículo "Motion Sickness in Public Road Transport: The Effect of Driver, Route, and Vehicle" (*Ergonomics*, 1999: 1646-1664) reportó datos sobre x = dosis de mareo provocado por el movimiento (calculado de acuerdo con una norma británica para evaluar movimientos similares en el mar) y y = náusea reportada (%). Las cantidades pertinentes son

$n = 17$, $\sum x_i = 222.1$, $\sum y_i = 193$, $\sum x_i^2 = 3056.69$, $\sum x_i y_i = 2759.6$, $\sum y_i^2 = 2975$

Los valores de dosis en la muestra oscilaron entre 6.0 y 17.6.

- **a.** Suponiendo que el modelo de regresión lineal simple es válido para relacionar estas dos variables (esto es apoyado por los datos sin procesar), calcule e interprete un estimador del parámetro de pendiente que dé información sobre la precisión y confiabilidad de la estimación.
- **b.** ¿Parece haber una relación lineal útil entre estas dos variables? Responda la pregunta empleando el método del valor P.
- **c.** ¿Sería sensible utilizar el modelo de regresión lineal simple como base para predecir el % de náusea cuando la dosis = 5.0? Explique su razonamiento.
- **d.** Cuando se utilizó Minitab para ajustar el modelo de regresión lineal simple a los datos sin procesar, la observación (6.0, 2.50) fue señalada como que posiblemente tiene un impacto sustancial en el ajuste. Elimine esta observación de la muestra y recalcule la estimación del inciso (a). Basado en esto, ¿parece ejercer la observación una influencia indebida?

36. Se produce una bruma (gotas transportadas por el aire o aerosoles) cuando se utilizan fluidos para remover metales en operaciones de maquinado para enfriar y lubricar la herramienta y la pieza de trabajo. La generación de bruma es una preocupación para la OSHA, la que recientemente ha reducido sustancialmente la norma del lugar de trabajo. El artículo "Variables Affecting Mist Generation from Metal Removal Fluids" (*Lubrication Engr.*, 2002: 10–17) dio los datos adjuntos sobre x = velocidad de flujo

de un aceite soluble al 5% (cm/s) y y = la cantidad de gotas de bruma con diámetro menor que 10 μm (mg/m^3):

x	89	177	189	354	362	442	965
y	.40	.60	.48	.66	.61	.69	.99

a. Los investigadores realizaron un análisis de regresión lineal simple para relacionar las dos variables. ¿Apoya la gráfica de dispersión esta estrategia?

b. ¿Qué proporción de la variación observada de la bruma puede ser atribuida a la relación de regresión lineal simple entre velocidad y bruma?

c. A los investigadores les interesaba particularmente el impacto en la bruma de la velocidad creciente de 100 a 1000 (un factor de 10 correspondiente a la diferencia entre los valores x más pequeños y más grandes presentes en la muestra). Cuando x se incrementa de esta manera, ¿existe evidencia sustancial de que el incremento promedio verdadero de y es menor que .6?

d. Estime el cambio promedio verdadero de la bruma asociado con un incremento de 1 cm/s en la velocidad y hágalo de modo que dé información sobre precisión y confiabilidad.

37. La obtención de imágenes por medio de resonancia magnética (MRI, por sus siglas en inglés) está bien establecida como una herramienta para medir velocidades de la sangre y flujos de volúmenes. El artículo "Correlation Analysis of Stenotic Aortic Valve Flow Patterns Using Phase Contrast MRI", citado en el ejercicio 1.67, propuso utilizar esta metodología para determinar el área valvular en pacientes con estenosis aórtica. Los datos adjuntos sobre velocidad pico (m/s) obtenidos de exámenes de 23 pacientes en dos planos diferentes se tomaron de una gráfica que aparece en el artículo citado.

Nivel-:	.60	.82	.85	.89	.95	1.01	1.01	1.05
Nivel--:	.50	.68	.76	.64	.68	.86	.79	1.03
Nivel-:	1.08	1.11	1.18	1.17	1.22	1.29	1.28	1.32
Nivel--:	.75	.90	.79	.86	.99	.80	1.10	1.15
Nivel-:	1.37	1.53	1.55	1.85	1.93	1.93	2.14	
Nivel--:	1.04	1.16	1.28	1.39	1.57	1.39	1.32	

a. ¿Parece haber alguna diferencia entre la velocidad promedio verdadera en los dos planos diferentes? Realice una prueba de hipótesis apropiada (como lo hicieron los autores del artículo).

b. Los autores del artículo también regresaron el nivel--velocidad contra nivel-velocidad. La intersección y la pendiente estimadas resultantes son .14701 y .65393 con errores estándar estimados correspondientes de .07877 y .05947, coeficiente de determinación de .852 y s = .110673. El artículo incluyó un comentario de que esta regresión mostraba evidencias de una fuerte relación lineal pero una pendiente de regresión muy por debajo de 1. ¿Está de acuerdo?

38. Remítase a los datos sobre x = tasa de liberación y y = tasa de emisión de NO_x dados en el ejercicio 19.

a. ¿Especifica el modelo de regresión lineal simple una relación útil entre las dos tasas? Use el procedimiento de prueba apropiado para obtener información sobre el valor P y luego saque una conclusión a nivel de significación de .01.

b. Calcule un intervalo de confianza de 95% para el cambio esperado en la tasa de emisiones asociado con un incremento de 10 MBtu/h-pie^2 en la tasa de liberación.

39. Realice la prueba de utilidad de modelo por medio del método ANOVA con los datos de contenido de humedad-tasa de filtración del ejemplo 12.6. Verifique que da una resultado equivalente al de la prueba t.

40. Use las reglas del valor esperado para demostrar que $\hat{\beta}_0$ es un estimador insesgado de β_0 (suponiendo que $\hat{\beta}_1$ es insesgado para β_1).

41. **a.** Verifique que $E(\hat{\beta}_1) = \beta_1$ con las reglas de valor esperado del capítulo 5.

b. Use las reglas de varianza del capítulo 5 para verificar la expresión para $V(\hat{\beta}_1)$ dada en esta sección.

42. Verifique que si cada x_i se multiplica por una constante positiva c y cada y_i se multiplica por otra constante positiva d, el estadístico t para probar H_0: $\beta_1 = 0$ contra H_a: $\beta_1 \neq 0$ no cambia de valor (el valor de $\hat{\beta}_1$ cambiará, lo que demuestra que la magnitud de $\hat{\beta}_1$ no es indicativo por sí mismo de la utilidad de modelo).

43. La probabilidad de un error de tipo II con la prueba t para H_0: $\beta_1 = \beta_{10}$ se calcula del mismo modo que para las pruebas t del capítulo 8. Si el valor alternativo de β_1, es denotado por β_1', el valor de

$$d = \frac{|\beta_{10} - \beta_1'|}{\sigma\sqrt{\dfrac{n-1}{\sum x_i^2 - (\sum x_i)^2/n}}}$$

se calcula primero, luego se ingresa al conjunto apropiado de curvas de la tabla A.17 del Apéndice por el eje horizontal con el valor de d y β se lee en la curva de $n - 2$ grados de libertad. Un artículo que apareció en el *Journal of Public Health Engineering* reporta los resultados de un análisis de regresión basado en n = 15 observaciones en las cuales x = temperatura de aplicación de filtro (°C) y y = % de eficiencia de eliminación de BOD. Las cantidades calculadas incluyen $\sum x_i$ = 402, $\sum x_i^2$ = 11,098, s = 3.725 y $\hat{\beta}_1$ = 1.7035. Considere probar a un nivel de .01 H_0: $\beta_1 = 1$, la que manifiesta que el incremento esperado en el % de eliminación de BOD es 1 cuando la temperatura de aplicación del filtro se incrementa 1°C, contra la alternativa H_a: $\beta_1 > 1$. Determine P (error de tipo II) cuando $\beta_1' = 2$, $\sigma = 4$.

12.4 Inferencias sobre $\mu_{Y\cdot x^*}$ y predicción de valores Y futuros

Sea x^* un valor específico de la variable independiente x. Una vez que las $\hat{\beta}_0$ y $\hat{\beta}_1$ estimadas han sido calculadas, $\hat{\beta}_0 + \hat{\beta}_1 x^*$ puede ser considerada como una estimación puntual de $\mu_{Y\cdot x^*}$ (el valor esperado o el valor promedio real de Y cuando $x = x^*$) o como una

predicción del valor Y que resultará de una sola observación realizada cuando $x = x^*$. La estimación puntual o predicción por sí misma no da información sobre qué tan precisamente $\mu_{Y \cdot x^*}$ ha sido estimada o Y ha sido pronosticada. Esto se remedia desarrollando un intervalo de confianza para $\mu_{Y \cdot x^*}$ y un intervalo de predicción (IP) para un solo valor de Y.

Antes de obtener datos muestrales, tanto $\hat{\beta}_0$ como $\hat{\beta}_1$ están sujetas a variabilidad de muestreo; es decir, ambos son estadísticos cuyos valores variarán de muestra en muestra. Supóngase, por ejemplo que $\beta_0 = 50$ y $\beta_1 = 2$. Entonces una primera muestra de pares (x, y) podría dar $\hat{\beta}_0 = 52.35$, $\hat{\beta}_1 = 1.895$, una segunda muestra podría dar $\hat{\beta}_0 = 46.52$, $\hat{\beta}_1 = 2.056$ y así sucesivamente. Se desprende que $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x^*$ misma cambia de valor de muestra en muestra, así que es un estadístico. Si la intersección y la pendiente de la línea de la población son los valores antes mencionados 50 y 2, respectivamente, y $x^* = 10$, entonces este estadístico está tratando de estimar el valor $50 + 2(10) = 70$. La estimación con una primera muestra podría ser $52.35 + 1.895(10) = 71.30$, con una segunda muestra podría ser $46.52 + 2.056(10) = 67.08$ y así sucesivamente.

Esta variación en el valor de $\hat{\beta}_0 + \hat{\beta}_1 x^*$ se puede visualizar regresando a la figura 12.13 en la página 492. Considere el valor $x^* = 300$. Las alturas de las 20 rectas de regresión estimada por encima de este valor son un poco diferentes entre sí. Lo mismo puede decirse de las alturas de las líneas por encima del valor $x^* = 350$. De hecho, parece que hay más variación en el valor de $\hat{\beta}_0 + \hat{\beta}_1(350)$ que en el de $\hat{\beta}_0 + \hat{\beta}_1(300)$. Veremos en breve que esto es porque 350 está más lejos de $\bar{x} = 235.71$ (el "centro de los datos") que 300.

Los métodos para hacer inferencias acerca de β_1 se basan en las propiedades de la distribución muestral del estadístico $\hat{\beta}_1$. De la misma manera, las inferencias sobre el valor medio Y de $\beta_0 + \beta_1 x^*$ se basan en las propiedades de la distribución muestral del estadístico $\hat{\beta}_0 + \hat{\beta}_1 x^*$. La sustitución de las expresiones para $\hat{\beta}_0$ y $\hat{\beta}_1$ en $\hat{\beta}_0 + \hat{\beta}_1 x^*$ y seguida de alguna manipulación algebraica lleva a la representación de $\hat{\beta}_0 + \hat{\beta}_1 x^*$ como una función lineal de las Y_i:

$$\hat{\beta}_0 + \hat{\beta}_1 x^* = \sum_{i=1}^{n} \left[\frac{1}{n} + \frac{(x^* - \bar{x})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \right] Y_i = \sum_{i=1}^{n} d_i Y_i$$

Los coeficientes $d_1, d_2, \ldots, d_n$ en esta función lineal implican las x_i y x^*, las cuales son fijas. La aplicación de las reglas de la sección 5.5 a esta función lineal da las siguientes propiedades.

PROPOSICIÓN

Sea $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x^*$, donde x^* es algún valor fijo de x. Entonces

1. El valor medio de $\hat{Y}$ es

$$E(\hat{Y}) = E(\hat{\beta}_0 + \hat{\beta}_1 x^*) = \mu_{\hat{\beta}_0 + \hat{\beta}_1 x^*} = \beta_0 + \beta_1 x^*$$

Así pues $\hat{\beta}_0 + \hat{\beta}_1 x^*$ es un estimador insesgado para $\beta_0 + \beta_1 x^*$ (es decir, de $\mu_{Y \cdot x^*}$).

2. La varianza de $\hat{Y}$ es

$$V(\hat{Y}) = \sigma_{\hat{Y}}^2 = \sigma^2 \left[\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum x_i^2 - (\sum x_i)^2/n} \right] = \sigma^2 \left[\frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}} \right]$$

y la desviación estándar $\sigma_{\hat{Y}}$ es la raíz cuadrada de esta expresión. La desviación estándar estimada de $\hat{\beta}_0 + \hat{\beta}_1 x^*$ denotada por $s_{\hat{Y}}$ o $s_{\hat{\beta}_0 + \hat{\beta}_1 x^*}$ se obtiene al reemplazar σ por su estimación s:

$$s_{\hat{Y}} = s_{\hat{\beta}_0+\hat{\beta}_1 x^*} = s\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}}$$

3. $\hat{Y}$ tiene una distribución normal.

La varianza de $\hat{\beta}_0 + \hat{\beta}_1 x^*$ es más pequeña cuando $x^* = \bar{x}$ y se incrementa a medida que x^* de aleja de $\bar{x}$ en una u otra dirección. Por consiguiente la estimación de $\mu_{Y\cdot x^*}$ es más precisa cuando x^* está cerca del centro de las x_i que cuando está lejos de los valores x a los cuales se realizaron las observaciones. Esto implicará tanto que el intervalo de confianza como el intervalo de predicción sean más angostos con una x^* cerca de $\bar{x}$ que con una x^* lejos de $\bar{x}$. La mayoría de los programas de computadora dan tanto $\hat{\beta}_0 + \hat{\beta}_1 x^*$ como $s_{\hat{\beta}_0+\hat{\beta}_1 x^*}$ con cualquier x^* especificado.

Inferencias sobre $\mu_{Y\cdot x^*}$

Así como los procedimientos inferenciales para β_1 se basaron en la variable t obtenida estandarizando β_1, una variable t obtenida estandarizando $\hat{\beta}_0 + \hat{\beta}_1 x^*$ conduce a un intervalo de confianza y procedimientos de prueba en este caso.

TEOREMA

La variable

$$T = \frac{\hat{\beta}_0 + \hat{\beta}_1 x^* - (\beta_0 + \beta_1 x^*)}{S_{\hat{\beta}_0+\hat{\beta}_1 x^*}} = \frac{\hat{Y} - (\beta_0 + \beta_1 x^*)}{S_{\hat{Y}}} \qquad (12.5)$$

tiene una distribución t con $n - 2$ grados de libertad.

Un enunciado de probabilidad implica que esta variable estandarizada ahora puede ser manipulada para producir un intervalo de confianza para μ_{Yx^*}

Un **intervalo de confianza** de $100(1 - \alpha)\%$ **para** $\mu_{Y\cdot x^*}$, el valor esperado de Y cuando $x = x^*$, es

$$\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\alpha/2,n-2} \cdot s_{\hat{\beta}_0+\hat{\beta}_1 x^*} = \hat{y} \pm t_{\alpha/2,\,n-2} \cdot s_{\hat{Y}} \qquad (12.6)$$

Este intervalo de confianza está centrado en la estimación puntual de μ_{Yx^*} y se extiende a cada lado en una cantidad que depende del nivel de confianza y del grado de variabilidad del estimador en el cual está basada la estimación puntual.

Ejemplo 12.13 La corrosión de varillas de refuerzo de acero es el problema de durabilidad más importante de estructuras de concreto reforzadas. La carbonatación del concreto ocurre a consecuencia de una reacción química que reduce el pH lo suficiente para iniciar la corrosión de las varillas de refuerzo. A continuación se dan datos representativos sobre x = profundidad de carbonatación (mm) y y = resistencia (MPa) para una muestra de especímenes testigo tomados de un edificio particular (tomados de una gráfica que aparece en el artículo “The Carbonation of Concrete Structures in the Tropical Environment of Singapore”, *Magazine of Concrete Res.*, 1996: 293–300).

x	8.0	15.0	16.5	20.0	20.0	27.5	30.0	30.0	35.0
y	22.8	27.2	23.7	17.1	21.5	18.6	16.1	23.4	13.4
x	38.0	40.0	45.0	50.0	50.0	55.0	55.0	59.0	65.0
y	19.5	12.4	13.2	11.4	10.3	14.1	9.7	12.0	6.8

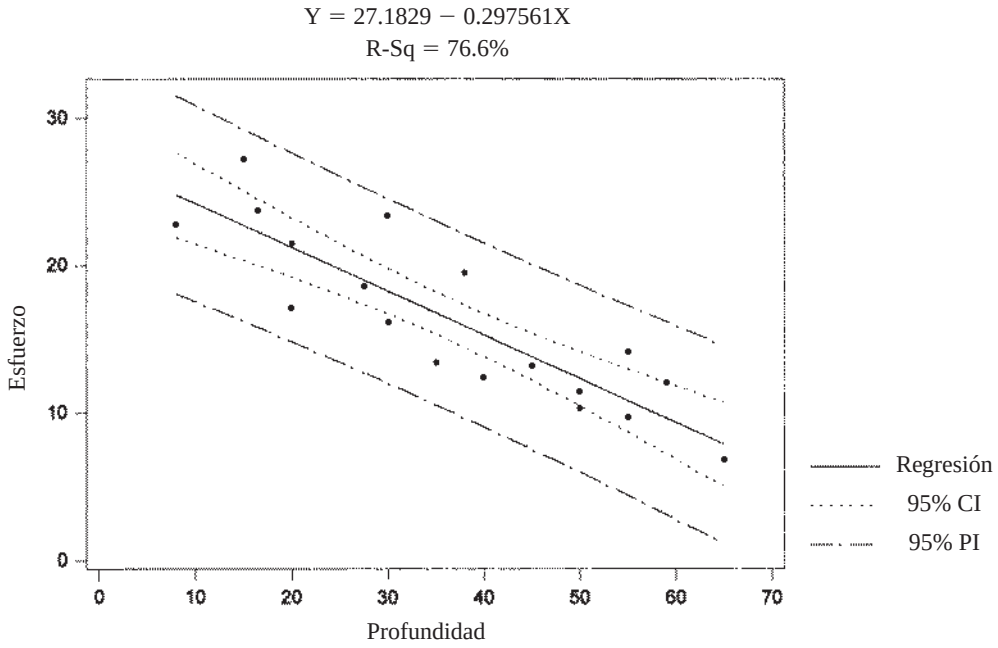


Figura 12.17 Diagrama de dispersión generado por Minitab con intervalos de confianza e intervalos de predicción con los datos del ejemplo 12.13

Una gráfica de dispersión de los datos (véase la figura 12.17) apoya fuertemente el uso del modelo de regresión lineal simple. Las siguientes son cantidades pertinentes:

$$\sum x_i = 659.0 \qquad \sum x_i^2 = 28{,}967.50 \qquad \bar{x} = 36.6111 \qquad S_{xx} = 4840.7778$$

$$\sum y_i = 293.2 \qquad \sum x_i y_i = 9293.95 \qquad \sum y_i^2 = 5335.76$$

$$\hat{\beta}_1 = -.297561 \qquad \hat{\beta}_0 = 27.182936 \qquad \text{SSE} = 131.2402$$

$$r^2 = .766 \qquad s = 2.8640$$

Calcúlese ahora un intervalo de confianza, utilizando un nivel de confianza de 95%, para la resistencia media de todos los especímenes testigo que tienen una profundidad de carbonatación de 45 mm; es decir, un intervalo de confianza para $\beta_0 + \beta_1(45)$. El intervalo está centrado en

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1(45) = 27.18 - .2976(45) = 13.79$$

La desviación estándar estimada del estadístico $\hat{Y}$ es

$$s_{\hat{Y}} = 2.8640\sqrt{\frac{1}{18} + \frac{(45 - 36.6111)^2}{4840.7778}} = .7582$$

El valor crítico t de 16 grados de libertad para un nivel de confianza de 95% es 2.120, con el cual se determina que el intervalo deseado es

$$13.79 \pm (2.120)(.7582) = 13.79 \pm 1.61 = (12.18, 15.40)$$

La angostura de este intervalo sugiere que se tiene información razonablemente precisa sobre el valor medio que se está estimando. Recuerde que si recalcula este intervalo para muestra tras muestra, a la larga aproximadamente el 95% de los intervalos calculados incluirían $\beta_0 + \beta_1(45)$. Sólo se puede esperar que este valor medio quede en el intervalo único que se calculó.

La figura 12.18 muestra resultados Minitab obtenidos por una solicitud de ajustar el modelo de regresión lineal simple y calcular intervalos de confianza para el valor medio de resistencia a profundidades de 45 mm y 35 mm. Los intervalos aparecen en la parte inferior de los resultados; obsérvese que el segundo intervalo es más angosto que el primero, porque 35 está mucho más cerca de $\bar{x}$ que 45. La figura 12.17 muestra (1) curvas correspondientes a los límites de confianza con cada valor x diferente y (2) límites de predicción que se discutirán en breve. Obsérvese cómo las curvas se alejan cada vez más a medida que x se aleja de $\bar{x}$.

```
La ecuación de regresión es resistencia = 27.2 − 0.298 profundidad

Predictor        Coef        Stdev      t-ratio       P
 Constante      27.183        1.651       16.46   0.000
   profundidad −0.29756      0.04116      −7.23   0.000
s = 2.864     R-sq = 76.6%     R-sq(adj) = 75.1%

Análisis de varianza

FUENTE          DF         SS         MS          F         P
Regresión        1     428.62     428.62      52.25     0.000
     Error      16     131.24       8.20
     Total      17     559.86

   Fit        DE.Fit           95.0% C.I.           95.0% P.I.
13.793         0.758     (12.185, 15.401)      (7.510, 20.075)

   Fit        DE.Fit           95.0% C.I.           95.0% P.I.
16.768         0.678     (15.330, 18.207)     (10.527, 23.009)
```

Figura 12.18 Resultados de regresión obtenidos con Minitab con los datos del ejemplo 12.13 ■

En algunas situaciones se desea un intervalo de confianza no sólo para un solo valor x sino para dos o más valores x. Supóngase que un investigador desea un intervalo de confianza tanto para $\mu_{Y \cdot v}$ como para $\mu_{Y \cdot w}$, donde v y w son dos valores diferentes de la variable independiente. Es tentador calcular el intervalo (12.6) primero con $x = v$ y luego con $x = w$. Supóngase que se utiliza $\alpha = .05$ en cada cálculo para obtener dos intervalos de 95%. Luego si las variables implicadas al calcular los dos intervalos fueran independientes una de otra, el nivel de confianza conjunto sería $(.95) \cdot (.95) \approx .90$.

Sin embargo, los intervalos no son independientes porque se utilizan las mismas $\hat{\beta}_0$, $\hat{\beta}_1$ y S en cada uno. Por consiguiente no se puede aseverar que el nivel de confianza conjunto para los dos intervalos sea exactamente de 90%. Se puede demostrar, no obstante, que si el intervalo de confianza de $100(1 - \alpha)\%$ (12.6) se calcula tanto con $x = v$ como con $x = w$ para obtener intervalos de confianza conjuntos para $\mu_{Y \cdot v}$ y $\mu_{Y \cdot w}$, entonces *el nivel de confianza conjunto en el par de intervalos resultante es por lo menos de* $100(1 - 2\alpha)\%$. En particular, si se utiliza $\alpha = .05$ se obtiene un nivel de confianza conjunto de *por lo menos* 90%, en tanto que si se utiliza $\alpha = .01$ se obtiene una confianza de por lo menos 98%. Así, en el ejemplo 12.13, un intervalo de confianza de 95% para $\mu_{Y \cdot 45}$ fue (12.185, 15.401) y un intervalo de confianza de 95% para μ_{Y35} fue (15.330, 18.207). El nivel de confianza simultáneo o conjunto para las dos proposiciones $12.185 < \mu_{Y \cdot 45} < 15.401$ y $15.330 < \mu_{Y \cdot 35} < 18.207$ es por lo menos de 90%.

La validez de estos intervalos de confianza conjuntos o simultáneos se fundamenta en un resultado de probabilidad llamado **desigualdad de Bonferroni,** así que los intervalos de confianza conjuntos se conocen como **intervalos de Bonferroni.** El método es fácil de generalizar para que dé intervalos conjuntos para k diferentes $\mu_{Y\cdot x}$. *Utilizando el intervalo* (12.6) *por separado primero con* $x = x_1^*$, *luego con* $x = x_2^*, \ldots,$ *y finalmente con* $x = x_k^*$ *se obtiene un conjunto de k intervalos de confianza con los cuales el nivel de confianza simultáneo o conjunto está garantizado que sea de por lo menos* $100(1 - k\alpha)\%$.

Las pruebas de hipótesis con respecto a $\beta_0 + \beta_1 x^*$ están basadas en el estadístico de prueba T obtenido reemplazando $\beta_0 + \beta_1 x^*$ en el numerador de (12.5) por el valor nulo de μ_0. Por ejemplo, H_0: $\beta_0 + \beta_1(45) = 15$ en el ejemplo 12.13 expresa que cuando la profundidad de carbonatación es de 45, la resistencia esperada (es decir, promedio verdadero) es de 15. El valor estadístico de prueba es entonces $t = [\hat{\beta}_0 + \hat{\beta}_1(45) - 15]/s_{\hat{\beta}_0 + \hat{\beta}_1(45)}$ y la prueba es de cola superior, inferior o de dos colas de acuerdo a la desigualdad en H_a.

Intervalo de predicción para un valor futuro de Y

A menos que se calcule un intervalo estimado para $\mu_{Y\cdot x^*}$ un investigador quizá desee obtener un intervalo de valores plausibles para el valor de Y asociado con alguna observación futura cuando la variable independiente tiene el valor x^*. Por ejemplo, el tamaño del vocabulario y está relacionado con la edad x de un niño. El intervalo de confianza (12.6) con $x^* = 6$ podría proporcionar un estimado del tamaño de vocabulario promedio verdadero de todos los niños de 6 años. Alternativamente, se podría desear un intervalo de valores plausibles para el tamaño del vocabulario de un niño particular de 6 años.

Un intervalo de confianza se refiere a un parámetro, o característica de población, cuyo valor es fijo pero desconocido. En contraste, un valor futuro de Y no es un parámetro sino una variable aleatoria; por eso se hace referencia a un intervalo de valores plausibles para un valor Y futuro como **intervalo de predicción** en lugar de intervalo de confianza. El error de estimación es $\beta_0 + \beta_1 x^* - (\hat{\beta}_0 + \hat{\beta}_1 x^*)$, una diferencia entre una cantidad fija (pero desconocida) y una variable aleatoria. El error de predicción es $Y - (\hat{\beta}_0 + \hat{\beta}_1 x^*)$, una diferencia entre dos variables aleatorias. Existe por lo tanto más incertidumbre en la predicción que en la estimación, así que un intervalo de predicción será más ancho que un intervalo de confianza. Como el valor futuro Y es independiente de las Y_i observadas,

$$\begin{aligned} V[Y - (\hat{\beta}_0 + \hat{\beta}_1 x^*)] &= \text{varianza del error de predicción} \\ &= V(Y) + V(\hat{\beta}_0 + \hat{\beta}_1 x^*) \\ &= \sigma^2 + \sigma^2\left[\frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}\right] \\ &= \sigma^2\left[1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}\right] \end{aligned}$$

Además, como $E(Y) = \beta_0 + \beta_1 x^*$ y $E(\hat{\beta}_0 + \hat{\beta}_1 x^*) = \beta_0 + \beta_1 x^*$, el valor esperado del error de predicción es $E(Y - (\hat{\beta}_0 + \hat{\beta}_1 x^*)) = 0$. Se puede demostrar entonces que la variable estandarizada

$$T = \frac{Y - (\hat{\beta}_0 + \hat{\beta}_1 x^*)}{S\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}}}$$

tiene una distribución t con $n - 2$ grados de libertad. Sustituyendo esta T en la proposición de probabilidad $P(-t_{\alpha/2,n-2} < T < t_{\alpha/2,n-1}) = 1 - \alpha$ y manipulándola para aislar Y entre las dos desigualdades se obtiene el siguiente intervalo.

Un **intervalo de predicción de** $100(1-\alpha)\%$ **para una observación** Y **futura que se va a realizar cuando** $x = x^*$ es

$$\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\alpha/2,n-2} \cdot s\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}}$$

$$= \hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\alpha/2,n-2} \cdot \sqrt{s^2 + s^2_{\hat{\beta}_0 + \hat{\beta}_1 x^*}} \qquad (12.7)$$

$$= \hat{y} \pm t_{\alpha/2,n-2} \cdot \sqrt{s^2 + s^2_{\hat{Y}}}$$

La interpretación del nivel de predicción de $100(1-\alpha)\%$ es idéntica al de los niveles de confianza previos; si se utiliza (12.7) repetidamente, a la larga los intervalos resultantes en realidad contendrán los valores y observados el $100(1-\alpha)\%$ del tiempo. Obsérvese que el 1 debajo de la raíz cuadrada inicial hace que el intervalo de predicción (12.7) sea más ancho que el intervalo de confianza (12.6), aun cuando ambos intervalos estén centrados en $\hat{\beta}_0 + \hat{\beta}_1 x^*$. Además, a medida que $n \to \infty$, el ancho del intervalo de confianza tiende a cero, en tanto que el ancho del intervalo de predicción no (porque incluso con el perfecto conocimiento de β_0 y β_1, seguirá habiendo incertidumbre en la predicción).

Ejemplo 12.14 Regrese a los datos de profundidad de carbonatación-resistencia del ejemplo 12.13 y calcule un intervalo de predicción de 95% para un valor de resistencia que resultaría de seleccionar un solo espécimen testigo cuya profundidad de carbonatación es de 45 mm. Cantidades pertinentes del ejemplo son

$$\hat{y} = 13.79 \quad s_{\hat{Y}} = .7582 \quad s = 2.8640$$

Para un nivel de predicción de 95% basado en $n - 2 = 16$ grados de libertad, el valor crítico es 2.120, exactamente el que se utilizó previamente para un nivel de confianza de 95%. El intervalo de predicción es entonces

$$13.79 \pm (2.120)\sqrt{(2.8640)^2 + (.7582)^2} = 13.79 \pm (2.120)(2.963)$$
$$= 13.79 \pm 6.28 = (7.51, 20.07)$$

Valores plausibles para una sola observación de resistencia cuando la profundidad es de 45 mm son (al nivel de predicción de 95%) entre 7.51 MPa y 20.07 MPa. El intervalo de confianza de 95% para una resistencia media cuando la profundidad es de 45 fue (12.18, 15.40). El intervalo de predicción es mucho más ancho que esto debido a los $(2.8640)^2$ extra bajo la raíz cuadrada. La figura 12.18, los resultados Minitab del ejemplo 12.13, muestran este intervalo así como también el intervalo de confianza. ■

La técnica Bonferroni puede ser empleada como en el caso del intervalo de confianza. Si se calcula un intervalo de predicción de $100(1-\alpha)\%$ para cada uno de k valores diferentes de x, el nivel de predicción simultánea o conjunta para los k intervalos es por lo menos de $100(1 - k\alpha)\%$.

EJERCICIOS Sección 12.4 (44–56)

44. El ajuste del modelo de regresión lineal simple a las $n = 27$ observaciones de x = módulo de elasticidad y y = resistencia a la flexión dados en el ejercicio 15 de la sección 12.2 dio por resultado $\hat{y} = 7.592$, $s_{\hat{Y}} = .179$ cuando $x = 40$ y $\hat{y} = 9.741$, $s_{\hat{Y}} = .253$ para $x = 60$.

a. Explique por qué $s_{\hat{Y}}$ es más grande cuando $x = 60$ que cuando $x = 40$.

b. Calcule un intervalo de confianza con un nivel de confianza de 95% para la resistencia promedio verdadera de todas las vigas cuyo módulo de elasticidad es de 40.

c. Calcule un intervalo de predicción con un nivel de predicción 95% para la resistencia de una sola viga cuyo módulo de elasticidad es 40.

d. Si se calcula un intervalo de confianza de 95% para la resistencia promedio verdadera cuando el módulo de elasticidad es

60, ¿cuál será el nivel de confianza simultáneo tanto para este intervalo como para el intervalo calculado en el inciso (b)?

45. Reconsidere los datos de contenido de humedad–tasa de filtración introducidos en el ejemplo 12.6 (véase también el ejemplo 12.7).

a. Calcule un intervalo de confianza de 90% para $\beta_0 + 125\beta_1$, el contenido de humedad promedio verdadera cuando la tasa de filtración es 125.

b. Pronostique el valor del contenido de humedad con un solo experimento en el cual la tasa de filtración es 125 utilizando un nivel de predicción de 90%. ¿Cómo se compara este intervalo al intervalo del inciso (a)? ¿Por qué es éste el caso?

c. ¿Cómo se compararían los intervalos de los incisos (a) y (b) con un intervalo de confianza y un intervalo de predicción cuando la tasa de filtración es 115? Responda sin calcular en realidad estos nuevos intervalos.

d. Interprete las hipótesis H_0: $\beta_0 + 125\beta_1 = 80$ y H_a: $\beta_0 + 125\beta_1 < 80$ y luego realice una prueba a un nivel de significación de .01.

46. La astringencia es la calidad de un vino que hace que la boca del bebedor lo sienta un poco áspero, seco y astringente. El documento "Analysis of Tannins in Red Wine Using Multiple Methods: Correlation with Perceived Astringency" (*Amer. J. of Enol. and Vitic.*, 2006: 481–485) informó sobre una investigación para evaluar la relación entre la percepción de la astringencia y la concentración de taninos utilizando diversos métodos analíticos. He aquí los datos proporcionados por los autores en x = concentración de taninos por la precipitación de proteínas y y = la astringencia percibida determinada por un panel de catadores.

x	.718	.808	.924	1.000	.667	.529	.514	.559
y	.428	.480	.493	.978	.318	.298	−.224	.198
x	.766	.470	.726	.762	.666	.562	.378	.779
y	.326	−.336	.765	.190	.066	−.221	−.898	.836
x	.674	.858	.406	.927	.311	.319	.518	.687
y	.126	.305	−.577	.779	−.707	−.610	−.648	−.145
x	.907	.638	.234	.781	.326	.433	.319	.238
y	1.007	−.090	−1.132	.538	−1.098	−.581	−.862	−.551

Las cantidades importantes se resumen como sigue:

$$\sum x_i = 19.404,\ \sum y_i = -.549,\ \sum x_i^2 = 13.248032,$$

$$\sum y_i^2 = 11.835795,\ \sum x_i y_i = 3.497811$$

$$S_{xx} = 13.248032 - (19.404)^2/32 = 1.48193150,$$

$$S_{yy} = 11.82637622$$

$$S_{xy} = 3.497811 - (19.404)(-.549)/32 = 3.83071088$$

a. Ajuste el modelo de regresión lineal simple a estos datos. Después, determine la proporción de la variación observada en la astringencia que se puede atribuir a la relación entre el modelo de astringencia y concentración de taninos.

b. Calcule e interprete un intervalo de confianza para la pendiente de la recta de regresión real.

c. Estime un promedio real de astringencia cuando la concentración de taninos es .6 y hágalo de una manera que transmita información acerca de la fiabilidad y precisión.

d. Estime la astringencia del vino de una muestra única cuya concentración de taninos es .6 y hágalo de una manera que transmita información acerca de la fiabilidad y precisión.

e. ¿Le parece que el promedio real de astringencia de una concentración de taninos de .7 es algo que no sea 0? Establezca y pruebe las hipótesis adecuadas.

47. El modelo de regresión lineal simple se ajusta muy bien a los datos de precipitación pluvial y volumen de escurrimiento dados en el ejercicio 16 de la sección 12.2. La ecuación de la recta de mínimos cuadrados es $y = -1.128 + .82697x$, $r^2 = .975$ y $s = 5.24$.

a. Use el hecho de que $s_{\hat{Y}} = 1.44$ cuando el volumen de la precipitación pluvial es de 40 m^3 para predecir el escurrimiento en una forma que transmita información sobre confiabilidad y precisión. ¿Sugiere el intervalo resultante que se dispone de información precisa sobre el valor de escurrimiento para esta futura observación? Explique su razonamiento.

b. Calcule un intervalo de predicción para escurrimiento cuando la precipitación pluvial es de 50 utilizando el mismo nivel de predicción del inciso (a). ¿Qué se puede decir sobre el nivel de predicción simultáneo para los dos intervalos que calculó?

48. El resumidero en un colector pluvial es la superficie de contacto entre el escurrimiento superficial y el conductor de desagüe. El inserto del resumidero es un dispositivo que mejora las propiedades supresoras de contaminantes de éste. El artículo "An Evaluation of the Urban Stormwater Pollutant Demoral Efficiency of Catch Basin Inserts" (*Water Envir. Res.*, 2005: 500–510) reportó pruebas de varios insertos en condiciones controladas en las que el flujo de entrada es muy parecido al que se puede esperar en el campo. Considere los siguientes datos, tomados de una gráfica que aparece en el artículo, para un tipo particular de inserto sobre x cantidad filtrada (1000s de litros) y y = % total de sólidos suspendidos eliminados.

x	23	45	68	91	114	136	159	182	205	228
y	53.3	26.9	54.8	33.8	29.9	8.2	17.2	12.2	3.2	11.1

Las cantidades resumidas son

$$\sum x_i = 1251,\ \sum x_i^2 = 199{,}365,\ \sum y_i = 250.6,$$

$$\sum y_i^2 = 9249.36,\ \sum x_i y_i = 21{,}904.4$$

a. ¿Avala la gráfica de dispersión la selección del modelo de regresión lineal simple? Explique.

b. Obtenga la ecuación de la recta de mínimos cuadrados.

c. ¿Qué proporción de la variación observada en el % de eliminación puede ser atribuida a la relación de modelo?

d. ¿Especifica el modelo de regresión lineal simple una relación útil? Realice una prueba de hipótesis apropiada con un nivel de significancia de .05.

e. ¿Existe una fuerte evidencia para concluir que por lo menos existe un 2% de reducción de la eliminación de sólidos suspendidos promedio verdadera asociada con un incremento

de 10,000 litros de la cantidad filtrada? Pruebe las hipótesis apropiadas con $\alpha = .05$.

f. Calcule e interprete un intervalo de confianza de 95% para el % eliminado promedio verdadero cuando la cantidad filtrada es de 100,000 litros. ¿Cómo se compara este intervalo en cuanto a ancho con el intervalo de confianza cuando la cantidad filtrada es de 200,000 litros?

g. Calcule e interprete un intervalo de predicción de 95% para % eliminado cuando la cantidad filtrada es de 100,000 litros. ¿Cómo se compara este intervalo en cuanto a ancho con el intervalo de confianza calculado en (f) y con intervalo de predicción cuando la cantidad filtrada es de 200,000 litros?

49. Le informan que un intervalo de confianza de 95% para el contenido de plomo esperado cuando el flujo de tráfico es de 15, basado en una muestra de $n = 10$ observaciones es (462.1, 597.7). Calcule un intervalo de confianza de 99% para el contenido de plomo esperado con un nivel de confianza de flujo de tráfico es de 15.

50. Se han utilizado aleaciones de silicio-germanio en ciertos tipos de celdas solares. El artículo "Silicon-Germanium Films Deposited by Low-Frequency Plasma-Enhanced Chemical Vapor Deposition" (*J. of Material Res.*, 2006: 88–104) reportó sobre un estudio de varias propiedades estructurales y eléctricas. Considere los datos adjuntos sobre x = concentración de Ge en fase sólida (desde 0 hasta 1) y y = posición de nivel Fermi (eV).

x	0	.42	.23	.33	.62	.60	.45	.87	.90	.79	1	1	1
y	.62	.53	.61	.59	.50	.55	.59	.31	.43	.46	.23	.22	.19

Un diagrama de dispersión muestra una relación lineal sustancial. He aquí una salida Minitab de un ajuste de mínimos cuadrados. [*Nota*: existen varias inconsistencias entre los datos dados en el artículo, la gráfica que allí aparece y la información resumida sobre un análisis de regresión.]

```
La ecuación de regresión es
Posición fermi nivel = 0.7217 − 0.4327 concentra-
ción de Ge

S = 0.0737573  R–Sq = 80.2%  R–Sq(adj) = 78.4%

Análisis de varianza

Fuente      GL        SS         MS      F      P
Regresión    1  0.241728  0.241728  44.43  0.000
Error       11  0.059842  0.005440
Total       12  0.301569
```

a. Obtenga una estimación de intervalo del cambio esperado en la posición del nivel Fermi asociado con un incremento de .1 en la concentración de Ge e interprete su estimación.

b. Obtenga una estimación de intervalo para la posición media del nivel Fermi cuando la concentración es de .50 e interprete su estimación.

c. Obtenga un intervalo de valores plausibles para la posición que resulta de una sola observación que ha de realizarse cuando la concentración es de .50, interprete su intervalo y compare con el intervalo de (b).

d. Obtenga intervalos de confianza simultáneos para la posición esperada cuando la concentración es de .3, .5 y .7; el nivel de confianza conjunto deberá ser por lo menos de 97%.

51. Remítase al ejemplo 12.12 en el cual x = velocidad en la pista de pruebas y y = velocidad de rodamiento de prueba.

a. Minitab dio $s_{\hat{\beta}_0+\hat{\beta}_1(45)} = .120$ y $s_{\hat{\beta}_0+\hat{\beta}_1(47)} = .186$. ¿Por qué la primera desviación estándar estimada es más pequeña que la segunda?

b. Use los resultados obtenidos con Minitab del ejemplo para calcular un intervalo de confianza de 95% para la velocidad de rodamiento de prueba esperada cuando la velocidad de prueba es = 45.

c. Use los resultados obtenidos con Minitab para calcular un intervalo de predicción de 95% para un solo valor de la velocidad de rodamiento cuando la velocidad de prueba es = 47.

52. El grabado con plasma es esencial en la transferencia de patrones de líneas finas en procesos de semiconductores de corriente. El artículo "Ion Beam-Assisted Etching of Aluminum with Chlorine " (*J. of the Electrochem. Soc.*, 1985: 2010–2012) da los datos adjuntos (tomados de una gráfica) sobre flujo de cloro (x, en SCCM) a través de una tobera utilizada en el mecanismo de grabado y en la velocidad de grabado (y, en 100 A/min).

x	1.5	1.5	2.0	2.5	2.5	3.0	3.5	3.5	4.0
y	23.0	24.5	25.0	30.0	33.5	40.0	40.5	47.0	49.0

Las cantidades resumidas son $\sum x_i = 24.0$, $\sum y_i = 312.5$, $\sum x_i^2 = 70.50$, $\sum x_i y_i = 902.25$, $\sum y_i^2 = 11{,}626.75$, $\hat{\beta}_0 = 6.448718$, $\hat{\beta}_1 = 10.602564$.

a. ¿Especifica el modelo de regresión lineal simple una relación útil entre el flujo de cloro y la velocidad de grabado?

b. Estime el cambio promedio verdadero en la velocidad de grabado asociado con un incremento de 1 SCCM en la velocidad de flujo utilizando un intervalo de confianza de 95% e interprete el intervalo.

c. Calcule un intervalo de confianza de 95% para $\mu_{Y \cdot 3.0}$, la velocidad de grabado promedio verdadera cuando el flujo = 3.0. ¿Ha sido estimado con precisión este promedio?

d. Calcule un intervalo de predicción de 95% para una sola observación futura de velocidad de grabado que se realizará cuando el flujo = 3.0. ¿Es probable que sea precisa la predición?

e. ¿Serían los intervalos de confianza y predicción de 95% cuando el flujo = 2.5 sea más ancho o más angosto que los intervalos correspondientes de los incisos (c) y (d)? Responda sin que en realidad calcule los intervalos.

f. ¿Recomendaría calcular un intervalo de predicción de 95% para un flujo de 6.0? Explique.

53. Considere los siguientes cuatro intervalos basados en los datos del ejercicio 12.17 (sección 12.2):

a. Un intervalo de confianza de 95% para la porosidad media cuando el peso unitario es de 110

b. Un intervalo de predicción de 95% para la porosidad cuando el peso unitario es de 110

c. Un intervalo de confianza de 95% para la porosidad media cuando el peso unitario es de 115

d. Un intervalo de predicción de 95% para la porosidad cuando el peso unitario es de 115

Sin calcular alguno de estos intervalos, ¿qué se puede decir sobre sus anchos uno con respecto al otro?

54. La declinación de los abastos de agua en ciertas áreas de Estados Unidos ha creado la necesidad de incrementar el conocimiento de las relaciones entre factores económicos tales como

rendimiento de cosechas y factores hidrológicos y de suelos. El artículo "Variability of Soil Water Properties and Crop Yield in a Sloped Watershed" (*Water Resources Bull.*, 1988: 281–288) da datos sobre cosechas de sorgo (y, en g/m-surco) y distancia pendiente arriba (x, en m) en una cuenca inclinada. En la tabla adjunta se dan observaciones seleccionadas.

x	0	10	20	30	45	50	70
y	500	590	410	470	450	480	510
x	80	100	120	140	160	170	190
y	450	360	400	300	410	280	350

a. Construya una gráfica de dispersión. ¿Parece ser plausible este modelo de regresión lineal simple?
b. Realice una prueba de la utilidad del modelo.
c. Estime el rendimiento promedio verdadero cuando la distancia pendiente arriba es de 75 dando un intervalo de valores plausibles.

55. Verifique que en realidad $V(\hat{\beta}_0 + \hat{\beta}_1 x)$ está dada por la expresión que aparece en el texto. [*Sugerencia*: $V(\sum d_i Y_i) = \sum d_i^2 \cdot V(Y_i)$.]

56. El artículo "Bone Density and Insertion Torque as Predictors of Anterior Cruciate Ligament Graft Fixation Strength" (*The Amer. J. of Sports Med.*, 2004: 1421–1429) dio los datos adjuntos sobre par de torsión de inserción máximo (N · m) y carga de cedencia (N), donde ésta mide la resistencia del injerto, correspondientes a 15 especímenes diferentes.

Torque	1.8	2.2	1.9	1.3	2.1	2.2	1.6	2.1
Carga	491	477	598	361	605	671	466	431
Torque	1.2	1.8	2.6	2.5	2.5	1.7	1.6	
Carga	384	422	554	577	642	348	446	

a. ¿Es plausible que la carga de cedencia esté normalmente distribuida?
b. Estime la carga de cedencia promedio verdadera calculando un intervalo de confianza con un nivel de confianza de 95% e interprételo.
c. Los siguientes son resultados obtenidos con Minitab para la regresión de la carga de cedencia generada por el momento de torsión. ¿Especifica el modelo de regresión lineal simple una relación útil entre las variables?

```
Predictor      Coef    DE Coef       T       P
Constant     152.44      91.17    1.67   0.118
Torque       178.23      45.97    3.88   0.002

S = 73.2141  R–Sq = 53.6%  R–Sq(adj) = 50.0%

Fuente          GL      SS     MS      F      P
Regresión        1   80554  80554  15.03  0.002
Error residual  13   69684   5360
Total           14  150238
```

d. Los autores del artículo citado expresan, "por consiguiente, no se puede sino concluir que los métodos basados en análisis de regresión simple no son clínicamente suficientes para predecir la resistencia de fijación individual". ¿Está de acuerdo? [*Sugerencia*: considere predecir la carga de cedencia cuando el momento de torsión es de 2.0.]

12.5 Correlación

Existen muchas situaciones en las que el objetivo al estudiar el comportamiento conjunto de dos variables es ver si están relacionadas, en lugar de utilizar una para predecir el valor de la otra. En esta sección, primero se desarrolla el coeficiente de correlación muestral r como una medida de qué tan fuerte es la relación entre dos variables x y y en una muestra y luego se relaciona r con el coeficiente de correlación ρ definido en el capítulo 5.

Coeficiente de correlación muestral r

Dados n pares numéricos $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, es natural hablar de que x y y tienen una relación positiva si las x grandes se aparean con y grandes y las x pequeñas con y pequeñas. Asimismo, si las x grandes se aparean con y pequeñas y las x pequeñas con y grandes, entonces se implica una relación negativa entre las variables. Considérese la cantidad

$$S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}$$

Entonces si la relación es fuertemente positiva, una x_i por encima de la media $\bar{x}$ tenderá a aparearse con una y_i por encima de la media $\bar{y}$, de modo que $(x_i - \bar{x})(y_i - \bar{y}) > 0$ y este producto también será positivo siempre que tanto x_i como y_i estén por debajo de sus medias respectivas. De este modo una relación positiva implica que S_{xy} será positiva. Un argumento análogo demuestra que cuando la relación es negativa, S_{xy} será negativa, puesto que la mayoría de los productos $(x_i - \bar{x})(y_i - \bar{y})$ seguirán siendo negativos. Esto se ilustra en la figura 12.19.

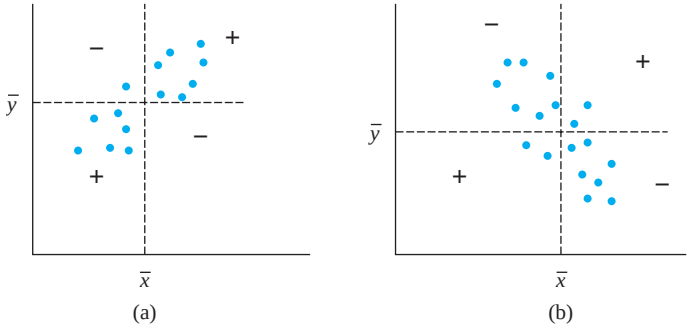


Figura 12.19 (a) Gráfica de dispersión con S_{xy} positiva; (b) gráfica de dispersión con S_{xy} negativa [+ significa $(x_i - \bar{x})(y_i - \bar{y}) > 0$, y − significa $(x_i - \bar{x})(y_i - \bar{y}) < 0$]

Aunque S_{xy} parece ser una medida plausible de la fuerza de una relación, aún no se sabe qué tan positiva o negativa pueda ser. Desafortunadamente, S_{xy} tiene un serio defecto. Si se cambian las unidades de medición de x o y, se puede hacer que S_{xy} sea arbitrariamente grande en magnitud o arbitrariamente próxima a cero. Por ejemplo, si $S_{xy} = 25$ cuando x se mide en metros, en ese caso $S_{xy} = 25{,}000$ cuando x se mide en milímetros y .025 cuando x está expresada en kilómetros. Una condición razonable para imponer cualquier medida de qué tan fuerte es la relación entre x y y es que la medida calculada no deberá depender de las unidades particulares utilizadas para medirlas. Esta condición se cumple modificando S_{xy} para obtener el coeficiente de correlación muestral.

DEFINICIÓN

El **coeficiente de correlación muestral** para los n pares $(x_1, y_1), \ldots, (x_n, y_n)$ es

$$r = \frac{S_{xy}}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}} = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}} \qquad (12.8)$$

Ejemplo 12.15 Una evaluación precisa de la productividad del suelo es crítica para una planificación racional del uso del suelo. Desafortunadamente, como el autor del artículo "Productivity Ratings Based on Soil Series" (*Prof. Geographer*, 1980: 158–163) argumenta, no es fácil obtener un índice de productividad del suelo aceptable. Una dificultad es que la productividad está determinada en parte por el tipo de cosecha y la relación entre el rendimiento de dos cosechas diferentes plantadas en el mismo suelo puede no ser muy fuerte. Como ilustración, el artículo presenta los datos adjuntos sobre una cosecha de maíz x y una cosecha de cacahuates y (mT/Ha) para ocho tipos diferentes de suelo.

x	2.4	3.4	4.6	3.7	2.2	3.3	4.0	2.1
y	1.33	2.12	1.80	1.65	2.00	1.76	2.11	1.63

Con $\sum x_i = 25.7$, $\sum y_i = 14.40$, $\sum x_i^2 = 88.31$, $\sum x_i y_i = 46.856$ y $\sum y_i^2 = 26.4324$,

$$S_{xx} = 88.31 - \frac{(25.7)^2}{8} = 5.75 \quad S_{yy} = 26.4324 - \frac{(14.40)^2}{8} = .5124$$

$$S_{xy} = 46.856 - \frac{(25.7)(14.40)}{8} = .5960$$

de donde

$$r = \frac{.5960}{\sqrt{5.75}\sqrt{.5124}} = .347$$

■

Propiedades de *r*

Las propiedades más importantes de r son las siguientes:

1. El valor de r no depende de cuál de las dos variables estudiadas es x y cuál es y.
2. El valor de r es independiente de las unidades en las cuales x y y estén medidas.
3. $-1 \le r \le 1$
4. $r = 1$ si y sólo si todos los pares (x_i, y_i) quedan en una línea recta con pendiente positiva, y $r = -1$ si y sólo si los pares (x_i, y_i) quedan en una línea recta con pendiente negativa.
5. El cuadrado del coeficiente de correlación muestral da el valor del coeficiente de determinación que resultaría de ajustar el modelo de regresión lineal simple–en símbolos $(r)^2 = r^2$.

La propiedad 1 contrasta con lo que sucede en el análisis de regresión donde virtualmente todas las cantidades de interés (la pendiente estimada, la intersección y estimada, s^2, etc.) dependen de cuál de las dos variables sea tratada como la variable dependiente. Sin embargo, la propiedad 5 demuestra que la proporción de variación de la variable dependiente explicada al ajustar el modelo de regresión lineal simple no depende de cuál variable desempeñe este rol.

La propiedad 2 equivale a decir que r no cambia si cada x_i es reemplazada por cx_i y si cada y_i es reemplazada por dy_i (un cambio en la escala de medición), así como también si cada x_i es reemplazada por $x_i - a$ y y_i por $y_i - b$ (lo que cambia la ubicación de cero en el eje de medición). Esto implica, por ejemplo, que r es el mismo si la temperatura se mide en °F o °C.

La propiedad 3 dice que el valor máximo de r, correspondiente al grado más grande posible de relación positiva, es $r = 1$, mientras que la relación más negativa está identificada con $r = -1$. De acuerdo con la propiedad 4, las correlaciones positivas y negativas más grandes se obtienen sólo cuando todos los puntos quedan a lo largo de una línea recta. Cualquier otra configuración de puntos, aun cuando la configuración sugiere una relación determinística entre las variables, dará un valor r menor que 1 en magnitud absoluta. Por consiguiente *r mide el grado de relación lineal* entre las variables. Un valor de r cercano a cero no es evidencia de la falta de una fuerte relación, sino sólo de la ausencia de una relación lineal, de modo que tal valor de r debe ser interpretado con precaución. La figura 12.20 ilustra varias configuraciones de puntos asociadas con valores diferentes de r.

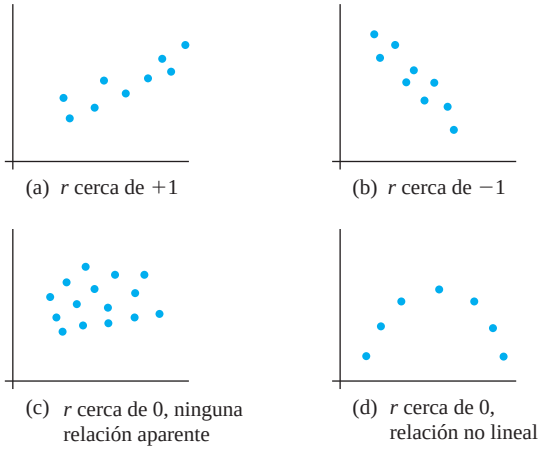


Figura 12.20 Gráficas de dispersión con valores diferentes de r

Una pregunta frecuentemente planteada es, "¿cuándo se puede decir que existe una correlación fuerte entre las variables y cuándo es débil?". He aquí una regla de oro informal para caracterizar el valor de r:

Débil	**Moderada**	**Fuerte**
$-.5 \le r \le .5$	si $-.8 < r < -.5$ o $.5 < r < .8$	si $r \ge .8$ o $r \le -.8$

Puede que le sorprenda que una r tan sustancial como .5 o $-.5$ vaya en la categoría débil. La razón es que si $r = .5$ o $-.5$, entonces $r^2 = .25$ en una regresión con cualquiera de las variables en el papel de y. Un modelo de regresión que explica la mayor parte del 25% de la variación observada en realidad no es muy impresionante. En el ejemplo 12.15, la correlación entre la cosecha de maíz y la cosecha de cacahuates se describiría como débil.

Inferencias sobre el coeficiente de correlación de una población

El coeficiente de correlación r mide qué tan fuerte es la relación entre x y y en la muestra observada. Se puede pensar que los pares (x_i, y_i) se sacaron de una población de pares bivariantes, con (X_i, Y_i) teniendo alguna función masa de probabilidad conjunta (X_i, Y_i) teniendo alguna. En el capítulo 5, el coeficiente de correlación $\rho(X, Y)$ se definió como

$$\rho = \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

donde

$$\text{Cov}(X, Y) = \begin{cases} \sum_x \sum_y (x - \mu_X)(y - \mu_Y) p(x, y) & (X, Y) \text{ discreto} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f(x, y)\, dx\, dy & (X, Y) \text{ continuo} \end{cases}$$

Si se considera que $p(x, y)$ o $f(x, y)$ describen la distribución de pares de valores dentro de toda la población, ρ se transforma en una medida de qué tan fuertemente están relacionadas x y y en la población. Propiedades de ρ análogas a aquellas para r se dieron en el capítulo 5.

El coeficiente de correlación de la población ρ es un parámetro o característica de la población, exactamente como lo son μ_X, μ_Y, σ_X y σ_Y, así que se puede utilizar el coeficiente de correlación muestral para hacer varias inferencias sobre ρ. En particular, r es una estimación puntual de ρ y el estimador correspondiente es

$$\hat{\rho} = R = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2} \sqrt{\sum (Y_i - \bar{Y})^2}}$$

Ejemplo 12.16 En algunos lugares, existe una fuerte asociación entre las concentraciones de dos contaminantes diferentes. El artículo "The Carbon Component of the Los Angeles Aerosol: Source Apportionment and Contributions to the Visibility Budget" *of* (*J. of Air Pollution Control Fed.*, 1984: 643-650) reporta los datos adjuntos sobre concentración de ozono x (ppm) y concentración de carbón secundario y (μg/m^3).

x	.066	.088	.120	.050	.162	.186	.057	.100
y	4.6	11.6	9.5	6.3	13.8	15.4	2.5	11.8
x	.112	.055	.154	.074	.111	.140	.071	.110
y	8.0	7.0	20.6	16.6	9.2	17.9	2.8	13.0

Las cantidades resumidas son $n = 16$, $\sum x_i = 1.656$, $\sum y_i = 170.6$, $\sum x_i^2 = .196912$, $\sum x_i y_i = 20.0397$ y $\sum y_i^2 = 2253.56$, de donde

$$r = \frac{20.0397 - (1.656)(170.6)/16}{\sqrt{.196912 - (1.656)^2/16}\sqrt{2253.56 - (170.6)^2/16}}$$

$$= \frac{2.3826}{(.1597)(20.8456)} = .716$$

La estimación puntual del coeficiente de correlación de la población ρ entre la concentración de ozono y la concentración de carbón secundaria es $\hat{\rho} = r = .716$. ■

Los intervalos de muestra pequeña y los procedimientos de prueba presentados en los capítulos 7–9 se basaron en la suposición de normalidad de la población. Para probar las hipótesis sobre ρ se debe hacer una suposición análoga sobre la distribución de los pares de valores (x, y) en la población. Ahora se supone que *tanto* X como Y son aleatorias, mientras que una gran parte del trabajo de regresión se realizó con x fijada por el experimentador :

SUPOSICIÓN

La distribución de probabilidad conjunta de (X, Y) está especificada por

$$f(x, y) = \frac{1}{2\pi \cdot \sigma_1\sigma_2\sqrt{1 - \rho^2}} e^{-[((x-\mu_1)/\sigma_1)^2 - 2\rho(x-\mu_1)(y-\mu_2)/\sigma_1\sigma_2 + ((y-\mu_2)/\sigma_2)^2]/[2(1-\rho^2)]}$$

$$-\infty < x < \infty$$

$$-\infty < y < \infty \qquad (12.9)$$

donde μ_1 y $\sigma_1$son la media y la desviación estándar de X, y μ_2 y σ_2 son la media y la desviación estándar de Y; $f(x, y)$ se denomina **distribución de probabilidad bivariante.**

La distribución normal bivariante es obviamente un tanto complicada, pero para los propósitos de este libro sólo se tiene que tener un conocimiento casual de varias de sus propiedades. La superficie determinada por $f(x, y)$ está por completo sobre el plano x, y [$f(x, y) \geq 0$] con apariencia de montículo o campana tridimensional, como se ilustra en la figura 12.21. Si se rebana la superficie con cualquier plano perpendicular al plano x, y, y se examina la curva dibujada en el "plano de corte", el resultado es una curva normal. Más precisamente, si $X = x$, se puede demostrar que la distribución (condicional) de Y es normal con media $\mu_{Y \cdot x} = \mu_2 - \rho\mu_1\sigma_2/\sigma_1 + \rho\sigma_2 x/\sigma_1$ y varianza $(1 - \rho^2)\sigma_2^2$. Éste es exactamente el modelo utilizado en la regresión lineal simple con $\beta_0 = \mu_2 - \rho\mu_1\sigma_2/\sigma_1$, $\beta_1 = \rho\sigma_2/\sigma_1$ y $\sigma^2 = (1 - \rho^2)\sigma_2^2$ independiente de x. La implicación es que *si los pares observados* (x_i, y_i) *en realidad se toman de una distribución normal bivariante, entonces el modelo de regresión lineal simple es una forma apropiada de estudiar el comportamiento de Y con x fija*. Si $\rho = 0$, entonces $\mu_{Y \cdot x} = \mu_2$ independiente de x; en realidad, cuando $\rho = 0$ la función de densidad de probabilidad conjunta $f(x, y)$ de (12.9) puede ser factorizada como $f_1(x)f_2(y)$, lo cual implica que X y Y son variables independientes.

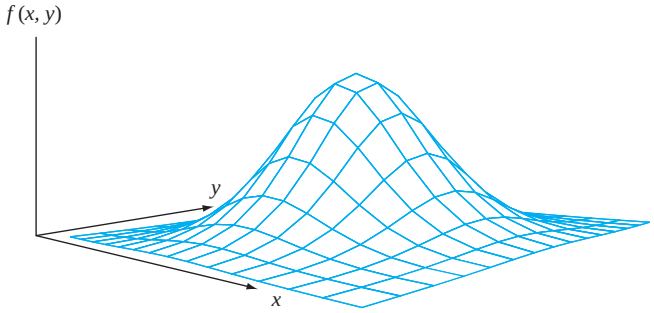


Figura 12.21 Gráfica de la función de densidad de probabilidad normal bivariante

Suponer que los pares se tomaron de una distribución normal bivariante permite probar hipótesis sobre ρ y construir un intervalo de confianza. No existe una forma completamente satisfactoria de verificar la plausibilidad de la suposición de normalidad bivariante. Una verificación parcial implica construir dos gráficas de probabilidad normal separadas, una para las x_i y otra para las y_i, puesto que la normalidad bivariante implica que las distribuciones marginales tanto de X como de Y son normales. Si cualquiera de las gráficas se aparta sustancialmente de un patrón de línea recta, no se deberán utilizar los siguientes procedimientos inferenciales para una n pequeña.

Prueba en cuanto a la ausencia de correlación

Cuando H_0: $\rho = 0$ es verdadera, el estadístico de prueba

$$T = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}}$$

tiene una distribución t con $n - 2$ grados de libertad.

Hipótesis alternativa	Región de rechazo para una prueba a nivel α
H_a: $\rho > 0$	$t \geq t_{\alpha,n-2}$
H_a: $\rho < 0$	$t \leq -t_{\alpha,n-2}$
H_a: $\rho \neq 0$	$t \geq t_{\alpha/2,n-2}$ o $t \leq -t_{\alpha/2,n-2}$

Un valor P basado en $n - 2$ grados de libertad puede ser calculado como previamente se describió.

Ejemplo 12.17 Los efectos neurotóxicos del manganeso son bien conocidos y normalmente son provocados por la prolongada exposición ocupacional durante largos lapsos de tiempo. En los campos de higiene ocupacional e higiene ambiental, la relación entre la peroxidación de lípidos, la cual es responsable del deterioro de los alimentos y de los daños de tejidos vivos, y la exposición ocupacional no ha sido previamente reportada. El artículo "Lipid Peroxidation in Workers Exposed to Manganese" (*Scand. J. of Work and Environ. Health*, 1996: 381–386) reportó datos sobre x = concentración de manganeso en sangre (ppb) y y = concentración (μmol/L) de malondialdehído, el cual es el producto estable de la peroxidación de lípidos, tanto para una muestra de 22 trabajadores expuestos a manganeso como para una muestra de control de 45 individuos. El valor de r para la muestra de control fue de .29, por lo que

$$t = \frac{(.29)\sqrt{45-2}}{\sqrt{1-(.29)^2}} \approx 2.0$$

El valor P correspondiente para una prueba t de dos colas basada en 43 grados de libertad es aproximadamente de .052 (el artículo citado reportó sólo que el valor $P > .05$). No se desearía rechazar la aseveración de que $\rho = 0$ al nivel de significación de .01 o .05. Para la muestra de trabajadores expuestos, $r = .83$ y $t \approx 6.7$, existe una clara evidencia de que existe una relación lineal en toda la población de trabajadores expuestos de la cual se seleccionó la muestra. ■

Como ρ mide al grado al cual existe una relación lineal entre las dos variables en la población, la hipótesis nula H_0: $\rho = 0$ manifiesta que no existe tal relación de población. En la sección 12.3 se utilizó la relación $t\ \hat{\beta}_1/s_{\hat{\beta}_1}$ para probar en cuanto a una relación lineal entre las dos variables en el contexto de análisis de regresión. Resulta que los dos procedimientos de prueba son completamente equivalentes porque $r\sqrt{n-2}/\sqrt{1-r^2} = \hat{\beta}_1/s_{\hat{\beta}_1}$. Cuando el interés radica sólo en valorar la fuerza de cualquier relación lineal en lugar de ajustarse a un modelo y utilizarlo para estimar o predecir, la fórmula del estadístico de prueba que se acaba de presentar requiere menos cálculos que la relación t.

Otras inferencias sobre ρ

El procedimiento para probar H_0: $\rho = \rho_0$ cuando $\rho_0 \neq 0$ no es equivalente a cualquier procedimiento de análisis de regresión. El estadístico de prueba se basa en una transformación de R llamada transformación de Fisher.

PROPOSICIÓN

Cuando $(X_1, Y_1), \ldots, (X_n, Y_n)$ es una muestra de una distribución normal bivariante, la variable aleatoria

$$V = \frac{1}{2}\ln\left(\frac{1+R}{1-R}\right) \tag{12.10}$$

tiene aproximadamente una distribución normal con media y varianza

$$\mu_V = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right) \qquad \sigma_V^2 = \frac{1}{n-3}$$

El razonamiento para la transformación es obtener una función de R que tenga una varianza independiente de ρ; éste no sería el caso con R misma. Además, no se deberá utilizar la transformación si n es bastante pequeña, puesto que la aproximación no será válida.

El estadístico de prueba para demostrar H_0: $\rho = \rho_0$ es

$$Z = \frac{V - \frac{1}{2}\ln[(1+\rho_0)/(1-\rho_0)]}{1/\sqrt{n-3}}$$

Hipótesis alternativa	Región de rechazo para una prueba a nivel α
H_a: $\rho > \rho_0$	$z \geq z_\alpha$
H_a: $\rho < \rho_0$	$z \leq -z_\alpha$
H_a: $\rho \neq \rho_0$	$z \geq z_{\alpha/2}$ o $z \leq -z_{\alpha/2}$

Se puede calcular un valor P del mismo modo que para pruebas z previas.

Ejemplo 12.18 El artículo "Size Effect in Shear Strength of Large Beams–Behavior and Finite Element Modelling" (*Mag. of Concrete Res.*, 2005: 497–509) reportó sobre un estudio de varias características de grandes vigas de alma profunda y poco profunda de concreto reforzado probadas hasta la falla. Considere los siguientes datos sobre x = resistencia de cubo y y = resistencia de cilindro (ambas en MPa):

x	55.10	44.83	46.32	51.10	49.89	45.20	48.18	46.70	54.31	41.50
y	49.10	31.20	32.80	42.60	42.50	32.70	36.21	40.40	37.42	30.80
x	47.50	52.00	52.25	50.86	51.66	54.77	57.06	57.84	55.22	
y	35.34	44.80	41.75	39.35	44.07	43.40	45.30	39.08	41.89	

Entonces $S_{xx} = 367.74$, $S_{yy} = 488.54$ y $S_{xy} = 322.37$, de donde $r = .761$. ¿Proporciona este valor una fuerte evidencia para concluir que las dos medidas de resistencia están por lo menos moderada y positivamente correlacionadas?

La interpretación previa de correlación positiva moderada fue $.5 < \rho < .8$, así que se desea probar H_0: $\rho = .5$ contra H_a: $\rho > .5$. El valor calculado de V es entonces

$$v = .5 \ln\left(\frac{1 + .761}{1 - .761}\right) = .999 \quad \text{y} \quad .5 \ln\left(\frac{1 + .5}{1 - .5}\right) = .549$$

Por consiguiente $z = (.999 - .549)\sqrt{19 - 3} = 1.80$. El valor P para una prueba de cola superior es .0359. La hipótesis nula por consiguiente puede ser rechazada a un nivel de significación de .05 pero no al nivel de .01. El último resultado es algo más sorprendente a la luz de la magnitud de r, pero cuando n es pequeño, puede resultar una r razonablemente grande aun cuando ρ no sea del todo sustancial. A nivel de significación de .01, la evidencia de una correlación moderadamente positiva no es convincente. ■

Para obtener un intervalo de confianza para ρ, primero se deduce un intervalo para $\mu_V = \frac{1}{2}\ln[(1 + \rho)/(1 - \rho)]$. Estandarizando V, escribiendo una proposición de probabilidad y manipulando las desigualdades resultantes se obtiene

$$\left(v - \frac{z_{\alpha/2}}{\sqrt{n - 3}}, v + \frac{z_{\alpha/2}}{\sqrt{n - 3}}\right) \tag{12.11}$$

como un intervalo de $100(1 - \alpha)\%$ para μ_V, donde $v = \frac{1}{2}\ln[(1 + r)/(1 - r)]$. Este intervalo puede entonces ser manipulado para dar un intervalo de confianza para ρ.

Un intervalo de confianza de $100(1 - \alpha)\%$ para ρ es

$$\left(\frac{e^{2c_1} - 1}{e^{2c_1} + 1}, \frac{e^{2c_2} - 1}{e^{2c_2} + 1}\right)$$

donde c_1 y c_2 son los puntos extremos izquierdo y derecho, respectivamente, del intervalo (12.11).

Ejemplo 12.19 El artículo "A Study of a Partial Nutrient Removal System for Wastewater Treatment Plants" (*Water Research*, 1972: 1389–1397) reporta sobre un método de eliminación de nitrógeno que implica el tratamiento del sobrenadante de un digestor aeróbico. Tanto el nitrógeno total afluente x (mg/L) como el porcentaje y de nitrógeno eliminado se registraron durante 20 días, con los siguientes estadísticos resultantes $\sum x_i = 285.90$, $\sum x_i^2 = 4409.55$, $\sum y_i = 690.30$, $\sum y_i^2 = 29{,}040.29$ y $\sum x_i y_i = 10{,}818.56$. El coeficiente de correlación muestral entre el nitrógeno afluente y el porcentaje de nitrógeno eliminado es $r = .733$ y se obtiene $v = .935$. Con $n = 20$, un intervalo de confianza de 95% para μ_V es $(.935 - 1.96/\sqrt{17}, .935 + 1.96/\sqrt{17}) = (.460, 1.410) = (c_1, c_2)$. El intervalo de 95% para ρ es

$$\left[\frac{e^{2(.46)} - 1}{e^{2(.46)} + 1}, \frac{e^{2(1.41)} - 1}{e^{2(1.41)} + 1}\right] = (.43, .89)$$ ■

En el capítulo 5, se advirtió que un valor grande del coeficiente de correlación (cercano a 1 o -1) implica sólo asociación y no causalidad. Esto es válido tanto para ρ como para r.

EJERCICIOS Sección 12.5 (57–67)

57. El artículo "Behavioral Effects of Mobile Telephone Use During Simulated Driving" (*Ergonomics*, 1995: 2536–2562) reportó que para una muestra de 20 sujetos experimentales, el coeficiente de correlación muestral con x = edad y y = tiempo desde que el sujeto obtuvo una licencia de manejo (años) fue .97. ¿Por qué piensa que el valor de r se aproxima tanto a 1? (Los autores del artículo dieron una explicación.)

58. El Turbine Oil Oxidation Test (TOST) y el Rotating Bomb Oxidation Test (RBOT) son dos procedimientos diferentes de evaluar la estabilidad ante la oxidación de aceites para turbina de vapor. El artículo "Dependence of Oxidation Stability of Steam Turbine Oil on Base Oil Composition" (*J. of the Society of Tribologists and Lubrication Engrs.*, octubre de 1997: 19–24) reportó las observaciones adjuntas sobre x = tiempo para realizar TOST (h) y y = tiempo para realizar RBOT (min) con 12 especímenes de aceite.

TOST	4200	3600	3750	3675	4050	2770
RBOT	370	340	375	310	350	200
TOST	4870	4500	3450	2700	3750	3300
RBOT	400	375	285	225	345	285

a. Calcule e interprete el valor del coeficiente de correlación muestral (como lo hicieron los autores del artículo).
b. ¿Cómo se vería afectado el valor de r si se hubiera hecho x = tiempo para realizar RBOT y y = tiempo para realizar TOST?
c. ¿Cómo se vería afectado el valor de r si el tiempo para realizar RBOT estuviera expresado en horas?
d. Construya gráficas de probabilidad normal y comente.
e. Realice una prueba de hipótesis para decidir si el tiempo para realizar RBOT y el tiempo para realizar TOST están linealmente relacionados.

59. La tenacidad y fibrosidad de los espárragos son importantes para determinar su calidad. Éste fue el enfoque de un estudio reportado en "Post-Harvest Glyphosphate Application Reduces Toughening, Fiber Content, and Lignification of Stores Asparagus Spears" (*J. of the Amer. Soc. of Hort. Science*, 1988: 569–572). El artículo reportó los datos adjuntos (tomados de una gráfica) sobre x = fuerza cortante (kg) y y = porcentaje de peso de fibra en seco.

x	46	48	55	57	60	72	81	85	94
y	2.18	2.10	2.13	2.28	2.34	2.53	2.28	2.62	2.63
x	109	121	132	137	148	149	184	185	187
y	2.50	2.66	2.79	2.80	3.01	2.98	3.34	3.49	3.26

$n = 18$, $\sum x_i = 1950$, $\sum x_i^2 = 251{,}970$,
$\sum y_i = 47.92$, $\sum y_i^2 = 130.6074$, $\sum x_i y_i = 5530.92$

a. Calcule el valor del coeficiente de correlación muestral. Basado en este valor, ¿cómo describiría la naturaleza de la relación entre las dos variables?
b. Si un primer espécimen tiene un valor más grande de fuerza cortante que un segundo espécimen, ¿qué tiende a ser cierto del porcentaje de peso de fibra en seco para los dos especímenes?
c. Si la fuerza cortante se expresa en libras, ¿qué le pasa al valor de r? ¿Por qué?
d. Si el modelo de regresión lineal simple fuera ajustado a estos datos, ¿qué proporción de la variación observada en porcentaje de peso de fibra en seco podría ser explicado por la relación de modelo?
e. Realice una prueba a un nivel de significación de .01 para decidir si existe una asociación lineal positiva entre las dos variables.

60. Las evaluaciones del movimiento de cabeza son importantes porque los individuos, especialmente los que son discapacita-

dos, pueden ser capaces de operar las comunicaciones de ayuda de esta manera. El artículo "Constancy of Head Turning Recorded in Healthy Young Humans" (*J. of Biomed. Engr.*, 2008: 428–436) reportó datos sobre los rangos en los ángulos de inclinación máxima de la cabeza en sentido de las manecillas del reloj para la parte anterior, posterior, derecha e izquierda en 14 sujetos seleccionados al azar. Considere los datos adjuntos para el ángulo promedio de inclinación máxima anterior (AMIA), tanto en la dirección de las manecillas del reloj (SH) y en sentido antihorario (SA).

Sujeto:	1	2	3	4	5	6	7
SH:	57.9	35.7	54.5	56.8	51.1	70.8	77.3
SA:	44.2	52.1	60.2	52.7	47.2	65.6	71.4

Sujeto:	8	9	10	11	12	13	14
SH:	51.6	54.7	63.6	59.2	59.2	55.8	38.5
SA:	48.8	53.1	66.3	59.8	47.5	64.5	34.5

a. Calcule una estimación puntual del coeficiente de correlación de población entre el SH y AMIA y la AMIA y SA ($\sum$SH = 786.7, $\sum$SA = 767.9, $\sum SH^2$ = 45,727.31, $\sum SA^2$ = 43,478.07, $\sum$SHSA = 44,187.87).

b. Suponiendo normalidad bivariante (gráficos de probabilidad normal del SH y muestras del SA son razonablemente rectas), lleve a cabo una prueba al nivel de significación .01 para decidir si existe una relación lineal entre las dos variables en la población (al igual que los autores del citado artículo). ¿La conclusión sería la misma si se utiliza un nivel de significancia de.001?

61. Los autores del artículo "Objective Effects of a Six Months' Endurance and Strength Training Program in Outpatients with Congestive Heart Failure" (*Medicine and Science in Sports and Exercise*, 1999: 1102-1107) presentaron un análisis de correlación para investigar la relación entre el nivel de lactato máximo x y la resistencia muscular y. Los datos adjuntos se tomaron de una gráfica incluida en el artículo.

x	400	750	770	800	850	1025	1200
y	3.80	4.00	4.90	5.20	4.00	3.50	6.30
x	1250	1300	1400	1475	1480	1505	2200
y	6.88	7.55	4.95	7.80	4.45	6.60	8.90

S_{xx} = 36.9839, S_{yy} = 2,628,930.357, S_{xy} = 7377.704. Un diagrama de dispersión muestra un patrón lineal.

a. Realice una prueba para ver si existe una correlación positiva entre el nivel de lactato máximo y la resistencia muscular en la población de la cual se seleccionaron estos datos.

b. Si se tuviera que realizar un análisis de regresión para predecir resistencia a consecuencia del nivel de lactato, ¿qué proporción de variación observada en la resistencia podría ser atribuida a la relación lineal aproximada? Responda la pregunta análoga si se utiliza regresión para predecir el nivel de lactato a partir de la resistencia y responda ambas preguntas sin que realice ningún cálculo de regresión.

62. Se conjetura que el contenido de hidrógeno es un factor importante en la porosidad de piezas fundidas de aleación de aluminio. El artículo "The Reduced Pressure Test as a Measuring Tool in the Evaluation of Porosity/Hydrogen Content in A1-7 Wt Pct Si-10 Vol. Pct SiC(p) Metal Matrix Composite" (*Metallurgical Trans.*, 1993: 1857–1868) da los datos adjuntos sobre x = contenido y y = porosidad al gas para una técnica de medición particular.

x	.18	.20	.21	.21	.21	.22	.23
y	.46	.70	.41	.45	.55	.44	.24
x	.23	.24	.24	.25	.28	.30	.37
y	.47	.22	.80	.88	.70	.72	.75

Minitab da los siguientes resultados en respuesta al comando Correlation:

```
Correlación de hidrógeno y porosidad = 0.449
```

a. Pruebe a un nivel de .05 para ver si el coeficiente de correlación de la población difiere de 0.

b. Si se hubiera realizado un análisis de regresión lineal simple, ¿qué porcentaje de la variación observada en la porosidad podría ser atribuido a la relación de modelo?

63. Se investigaron las propiedades físicas de seis muestras de tela retardante a las llamas en el artículo "Sensory and Physical Properties of Inherently Flame-Retardant Fabrics" (*Textile Research*, 1984: 61–68. Use los datos adjuntos y un nivel de significación de .05 para determinar si existe una relación lineal entre la rigidez x (mg-cm) y espesor y (mm). ¿Es sorprendente el resultado de la prueba a la luz del valor de r?

x	7.98	24.52	12.47	6.92	24.11	35.71
y	.28	.65	.32	.27	.81	.57

64. El artículo "Increases in Steroid Binding Globulins Induced by Tamoxifen in Patients with Carcinoma of the Breast" (*J. of Endocrinology*, 1978: 219–226) reporta datos sobre los efectos de la droga tamoxifeno en el cambio del nivel de globulina afín al cortisol (CBG, por sus siglas en inglés, *cortisol-binding globulin*) de pacientes durante el tratamiento. Con edad = x y ΔCBG = y, los valores resumidos son n = 26, $\sum x_i$ = 1613, $\sum(x_i - \bar{x})^2$ = 3756.96, $\sum y_i$ = 281.9, $\sum(y_i - \bar{y})^2$ = 465.34, y $\sum x_i y_i$ = 16,731.

a. Calcule un intervalo de confianza de 90% para el coeficiente de correlación verdadero ρ.

b. Pruebe H_0: $\rho = -.5$ contra H_a: $\rho < -.5$ al nivel de .05.

c. En un análisis de regresión de y en relación con x, ¿qué proporción de la variación del cambio del nivel de globulina afín al cortisol podría ser explicada por la variación de la edad del paciente dentro de la muestra?

d. Si decide realizar un análisis de regresión con la edad como variable dependiente, ¿qué proporción de la variación de la edad es explicable por la variación del ΔCBG?

65. La torsión durante la rotación externa de la cadera y la extensión pueden explicar por qué ocurren las lágrimas labral acetabular en atletas profesionales. El artículo "Hip Rotational Velocities During the Full Golf Swing" (*J. of Sports Science and Med.*, 2009: 296–299) informó sobre una investigación en la que el

pico de la velocidad máxima de rotación interna de la cadera (x) y el pico final de la velocidad de rotación externa de la cadera (y) se han determinado en una muestra de 15 jugadores de golf. Los datos proporcionados por los autores del artículo se utilizaron para calcular las siguientes cantidades resumidas:

$$\sum(x_i - \bar{x})^2 = 64{,}732.83, \sum(y_i - \bar{y})^2 = 130{,}566.96,$$

$$\sum(x_i - \bar{x})(y_i - \bar{y}) = 44{,}185.87$$

Gráficas separadas de probabilidad normal mostraron patrones lineales muy importantes.

a. Calcule una estimación puntual del coeficiente de correlación de la población.

b. Lleve a cabo una prueba al nivel de significación .01 para decidir si existe una relación lineal entre las dos velocidades en la población muestreada, su conclusión debe basarse en un valor P.

c. ¿La conclusión de (b) podría haber cambiado si hubiera probado la hipótesis adecuada para decidir si existe una relación lineal positiva entre la población? ¿Qué pasa si se utiliza un nivel de significación de .05 en lugar de .01?

66. Considere una serie de tiempo; es decir, una secuencia de observaciones $X_1, X_2, \ldots$ obtenidas durante el transcurso del tiempo con valores observados $x_1, x_2, \ldots, x_n$. Suponga que la serie no muestra tendencia hacia arriba o hacia abajo durante el transcurso del tiempo. Un investigador con frecuencia deseará saber qué tan fuertemente están relacionados los valores en la serie separados por un número especificado de unidades de tiempo. El *coeficiente de autocorrelación muestral correspondiente a un retardo* r_1 es simplemente el valor del coeficiente de correlación muestral r de los pares $(x_1, x_2), (x_2, x_3), \ldots, (x_{n-1}, x_n)$, es decir, pares de valores separados por una unidad de tiempo. Asimismo, el *coeficiente de autocorrelación muestral correspondiente a dos retardos* r_2 es r para los $n - 2$ pares (x_1, x_3), $(x_2, x_4), \ldots, (x_{n-2}, x_n)$.

a. Calcule los valores de r_1, r_2 y r_3 para los datos de temperatura del ejercicio 82 del capítulo 1 y comente.

b. Análogo al coeficiente de correlación de la población ρ, sean $\rho_1, \rho_2, \ldots$ los coeficientes de autocorrelación teóricos o de largo plazo con los varios retardos. Si todos estos ρ son 0, no existe relación (lineal) con cualquier retraso. En este caso, si n es grande, cada R_i tiene aproximadamente una distribución normal con media 0 y desviación estándar $1/\sqrt{n}$, y los R_i diferentes son casi independientes. Por consiguiente H_0: $\rho_i = 0$ puede ser rechazada a un nivel de significancia de aproximadamente .05 si $r_i \geq 2/\sqrt{n}$ o $r_i \leq -2/\sqrt{n}$. Si $n = 100$ y $r_1 = .16$, $r_2 = -.09$ y $r_3 = -.15$, ¿existe alguna evidencia de autocorrelación teórica con los primeros tres retrasos?

c. Si prueba simultáneamente la hipótesis nula del inciso (b) con más de un retraso, ¿por qué podría desear incrementar la constante de corte 2 en la región de rechazo?

67. Se recopiló una muestra de $n = 500$ pares (x,y) y se realizó una prueba de H_0: $\rho = 0$ contra H_a: $\rho \neq 0$. El valor P resultante se calculó como .00032.

a. ¿Qué conclusión sería apropiada a nivel de significación de .001?

b. ¿Indica este pequeño valor P que existe una relación muy fuerte entre x y y (un valor de ρ que difiera considerablemente de 0)? Explique.

c. Suponga ahora que una muestra de $n = 10{,}000$ pares (x, y) dio por resultado $r = .022$. Pruebe H_0: $\rho = 0$ contra H_a: $\rho \neq 0$ a un nivel de .05. ¿Es el resultado estadísticamente significativo? Comente sobre la significación práctica de su análisis.

EJERCICIOS SUPLEMENTARIOS (68–87)

68. El avalúo de un almacén puede parecer sencillo en comparación con otras asignaciones de avalúo. El avalúo de un almacén implica comparar una edificación que es principalmente una armadura abierta con otros edificios semejantes. Sin embargo, siguen habiendo varios atributos de un almacén que están plausiblemente relacionados con el valor apreciado. El artículo "Challenges In Appraising 'Simple' Warehouse Properties" (Donald Sonneman, *The Appraisal Journal*, abril de 2001, 174–178) dio los datos adjuntos sobre la altura de la armadura (pies), la cual determina qué tan alto pueden ser apilados los productos almacenados y el precio de venta ($) por pie cuadrado.

Altura:	12	14	14	15	15	16	18	22	22	24
Precio:	35.53	37.82	36.90	40.00	38.00	37.50	41.00	48.50	47.00	47.50

Altura de la armadura:	24	26	26	27	28	30	30	33	36
Precio de venta:	46.20	50.35	49.13	48.07	50.90	54.78	54.32	57.17	57.45

a. ¿Es el caso que la altura de la armadura y el precio de venta están "determinísticamente" relacionados, es decir, que el precio de venta está determinado por completo y únicamente por la altura de la armadura? [*Sugerencia*: Examine los datos.]

b. Construya una gráfica de dispersión de los datos. ¿Qué sugieren?

c. Determine la ecuación de la recta de mínimos cuadrados.

d. Dé una predicción puntual del precio cuando la altura de la armadura es de 27 pies y calcule el residuo correspondiente.

e. ¿Qué porcentaje de la variación observada del precio de venta puede ser atribuido a la relación lineal aproximada entre la altura de la armadura y el precio?

69. Remítase al ejercicio previo, el cual dio datos sobre alturas de armadura para una muestra de almacenes y los precios de venta correspondientes.

a. Estime el cambio promedio verdadero del precio de venta asociado con un pie de incremento de la altura de la armadura y hágalo de modo que dé información sobre la precisión de la estimación.

b. Estime el precio de venta promedio real de todos los almacenes cuya altura de armadura es de 25 pies y hágalo de modo que dé información sobre la precisión de la estimación.

c. Pronostique el precio de venta de un solo almacén cuya altura de armadura es de 25 pies y hágalo de modo que dé información sobre la precisión de la predicción. ¿Cómo se compara esta estimación con la estimación de (b)?

d. Sin calcular ningunos intervalos, ¿cómo se compararía el ancho de un intervalo de predicción de 95% con el precio de venta cuando la altura de la armadura es de 25 pies con el ancho de un intervalo de 95% cuando la altura es de 30 pies? Explique su razonamiento.

e. Calcule e interprete el coeficiente de correlación muestral.

70. A los científicos forenses con frecuencia les interesa realizar alguna clase de medición en un cuerpo (vivo o muerto) y luego utilizarla como base para inferir algo sobre la edad del cuerpo. Considere los datos adjuntos sobre edad (años) y % de ácido aspértico D (de aquí en adelante %DAA) de una pieza dental particular ("An Improved Method for Age at Death Determination from the Measurements of D-Aspertic Acid in Dental Collagen", *Archaeometry*, 1990: 61–70.)

Edad:	9	10	11	12	13	14	33	39	52	65	69
%DAA:	1.13	1.10	1.11	1.10	1.24	1.31	2.25	2.54	2.93	3.40	4.55

Suponga que una pieza dental de otro individuo tiene 2.01%DAA. ¿Podría ser el caso que el individuo tenga menos de 22 años? Esta pregunta era pertinente en cuanto a si el individuo podía ser sentenciado a cadena perpetua por homicidio o no.

Una estrategia aparentemente sensible es regresar la edad en % DDA, y luego calcule un intervalo de probabilidad para la edad cuando a %DAA = 2.01. No obstante, es más natural en este caso considerar la edad como la variable independiente x y el %DAA como la variable dependiente y, así que el modelo de regresión es $\%DAA = \beta_0 + \beta_1 x + \epsilon$. Después de estimar los coeficientes de regresión, se puede sustituir $y^* = 2.01$ en la ecuación estimada y luego resolverla para una predicción de edad $\hat{x}$. Este uso "inverso" de la línea de regresión se llama "calibración". Un intervalo de predicción para edad con nivel de predicción aproximadamente de $100(1 - \alpha)\%$ es $\hat{x} \pm t_{\alpha/2,n-2} \cdot SE$ donde

$$SE = \frac{s}{\hat{\beta}_1}\left\{1 + \frac{1}{n} + \frac{(\hat{x} - \bar{x})^2}{S_{xx}}\right\}^{1/2}$$

Calcule este intervalo de predicción para $y^* = 2.01$ y luego aborde la pregunta previamente planteada.

71. Los datos adjuntos sobre x = tasa de consumo de diesel medida por el método pesaje de drenaje y y = tasa medida por el método de trazado de intervalo de confianza, ambos en g/h, se tomaron de una gráfica incluida en el artículo "A New Measurement Method of Diesel Engine Oil Consumption Rate" (*J. of Soc. of Auto Engr*., 1985: 28–33).

x	4	5	8	11	12	16	17	20	22	28	30	31	39
y	5	7	10	10	14	15	13	25	20	24	31	28	39

a. Suponiendo que x y y están relacionadas por el modelo de regresión lineal simple, realice una prueba para decidir si es plausible que en promedio el cambio de la tasa medida por el método de trazado de intervalo de confianza sea idéntico al cambio de la tasa medido mediante el método de pesaje de drenaje.

b. Calcule e interprete el valor del coeficiente de correlación muestral.

72. Los resultados SAS dados al final de la página están basados en datos tomados del artículo "Evidence for and the Rate of Denitrification in the Arabian Sea" (*Deep Sea Research*, 1978: 431–435). Las variables estudiadas son x = nivel de salinidad (%) y y = nivel de nitrato (μM/L).

a. ¿Cuál es el tamaño de muestra n? [*Sugerencia*: busque los grados de libertad para SSE.]

b. Calcule una estimación puntual del nivel de nitrato esperado cuando el nivel de salinidad es de 35.5.

c. ¿Parece haber una relación lineal útil entre las dos variables?

d. ¿Cuál es el valor del coeficiente de correlación muestral?

e. ¿Utilizaría el modelo de regresión lineal simple para sacar conclusiones cuando el nivel de salinidad es de 40?

73. La presencia de carburos de aleación duros en aleaciones de hierro blanco al alto cromo produce una excelente resistencia a la abrasión, lo que las hace apropiadas para el manejo de materiales en las industrias minera y de procesamiento de materiales. Los datos adjuntos sobre x = contenido de austenita retenido (%) y y = pérdida por desgaste abrasivo (mm^3) en prueba de desgaste de alfileres con granate como el abrasivo se tomaron de una gráfica que aparece en el artículo "Microstructure-Property Relationships in High Chromium White Iron Alloys" (*Intl. Materials Reviews*, 1996: 59-82).

Resultados obtenidos con SAS para el ejercicio 72

```
Variable dependente : NITRLVL

                              Análisis de varianza

 Fuente          GL      Suma de cuadrados     Media cuadrática          Valor F        Prob > F

Modelo            1              64.49622            64.49622           63.309          0.0002
Error             6               6.11253             1.01875
C Total           7              70.60875

              Raíz MSE         1.00933      R-raíz        0.9134
              Dep Media       26.91250      Adj R-sq      0.8990
              C.V.             3.75043

                           Parámetros estimados

                          Parámetro             Error          T para HO:
Variable          GL        estimado           estándar       Parámetro = 0        Prob > |T|
INTERCEP           1      326.976038         37.71380243               8.670            0.0001
SALINIDAD          1       −8.403964          1.05621381              −7.957            0.0002
```

Resultados obtenidos con SAS para el ejercicio 73

```
Variable dependiente: ABRLOSS
                                  Análisis de varianza

 Fuente          GL      Suma de cuadrados     Media cuadrática          Valor F          Prob > F
  Modelo          1              0.63690              0.63690           15.444           0.0013
  Error          15              0.61860              0.04124
C Total          16              1.25551

                Raíz MSE       0.20308       R-square      0.5073
                Dep Media      1.10765       Adj R-sq      0.4744
                C.V.          18.33410

                                  Parámetros estimados

                          Parámetro              Error          T para H0:
Variable         GL        estimado           estándar       Parámetro = 0          Prob > |T|
INTERCEP          1        0.787218         0.09525879              8.264              0.0001
AUSTCONT          1        0.007570         0.00192626              3.930              0.0013
```

x	4.6	17.0	17.4	18.0	18.5	22.4	26.5	30.0	34.0
y	.66	.92	1.45	1.03	.70	.73	1.20	.80	.91
x	38.8	48.2	63.5	65.8	73.9	77.2	79.8	84.0	
y	1.19	1.15	1.12	1.37	1.45	1.50	1.36	1.29	

Use los datos y los resultados obtenidos con SAS dados al inicio de la página para responder las siguientes preguntas.

a. ¿Qué proporción de la variación observada de pérdida por desgaste puede ser atribuida a la relación de modelo de regresión lineal simple?

b. ¿Cuál es el valor del coeficiente de correlación muestral?

c. Pruebe la utilidad del modelo de regresión lineal simple con $\alpha = .01$.

d. Estime la pérdida por desgaste promedio verdadera cuando el contenido es de 50% y hágalo de modo que dé información sobre confiabilidad y precisión.

e. ¿Qué valor de pérdida por desgaste pronosticaría cuando el contenido es de 30% y cuál es valor del residuo correspondiente?

74. Los datos adjuntos se leyeron en una gráfica de dispersión del artículo "Urban Emissions Measured with Aircraft" (*J. of the Air and Waste Mgmt. Assoc.*, 1998: 16–25). La variable de respuesta es ΔNO_y y la variable explicativa es ΔCO.

ΔCO	50	60	95	108	135
ΔNO_y	2.3	4.5	4.0	3.7	8.2
ΔCO	210	214	315	720	
ΔNO_y	5.4	7.2	13.8	32.1	

a. Adapte un modelo apropiado a los datos y juzgue la utilidad del modelo.

b. Pronostique el valor de ΔNO_y que se obtendría al realizar una observación más cuando ΔCO es 400 y hágalo de modo que dé información sobre precisión y confiabilidad. ¿Parece que el ΔNO_y puede ser pronosticado con precisión? Explique.

c. El valor más grande de ΔCO es mucho más grande que los demás valores. ¿Ha tenido esta observación un impacto sustancial en la ecuación ajustada?

75. Se estudió la relación entre la velocidad (pies/s) y la cadencia al correr (número de pasos/s) entre corredoras de maratón. Las cantidades resumidas resultantes fueron $n = 11$, $\sum$(velocidad) = 205.4, $\sum$(velocidad)^2 = 3880.08, $\sum$(cadencia) = 35.16, $\sum$(cadencia)^2 = 112.681 y $\sum$(velocidad)(cadencia) = 660.130.

a. Calcule la ecuación de la recta de mínimos cuadrados que utilizaría para predecir la cadencia a partir de la velocidad.

b. Calcule la ecuación de la recta de mínimos cuadrados que utilizaría para predecir la velocidad a partir de la cadencia.

c. Calcule el coeficiente de determinación para la regresión de la cadencia basada en la velocidad del inciso (a) y para la regresión de la velocidad basada en la cadencia del inciso (b). ¿Cómo están relacionadas?

76. "Mezclabilidad de modo" se refiere a cuánto de la propagación de grietas es atribuible a los tres modos de fractura convencionales de abertura, deslizamiento o desgarro. Para problemas de aviones, sólo los dos primeros modos están presentes y el ángulo de mezclabilidad de modos mide el grado al cual la propagación se debe a deslizamiento en oposición a abertura. El artículo "Increasing Allowable Flight Loads by Improved Structural Modeling" (*AIAA J.*, 2006: 376–381] dio los siguientes datos sobre x = ángulo de mezclabilidad de modos (grados) y y = tenacidad a la fractura (N/m) de paneles utilizados en la construcción de aviones.

x	16.52	17.53	18.05	18.50	22.39	23.89	25.50	24.89
y	609.4	443.1	577.9	628.7	565.7	711.0	863.4	956.2
x	23.48	24.98	25.55	25.90	22.65	23.69	24.15	24.54
y	679.5	707.5	767.1	817.8	702.3	903.7	964.9	1047.3

a. Obtenga la ecuación de la línea de regresión estimada y discuta el grado al cual el modelo de regresión lineal simple es una forma razonable de relacionar la tenacidad a la fractura con el ángulo de modo de mezclabilidad.

b. ¿Sugieren los datos que el cambio promedio de la tenacidad a la fractura asociado con un incremento de un grado del ángulo de mezclabilidad de modos excede de 50 N/m? Realice una prueba apropiada de hipótesis.

c. Para propósitos de estimación con precisión de la pendiente de la línea de regresión de la población, ¿hubiera sido preferible realizar observaciones a los ángulos 16, 16, 18, 18, 20, 20, 20, 20, 22, 22, 22, 22, 24, 24, 26 y 26 (de nuevo un tamaño de muestra de 16)? Explique su razonamiento.

d. Calcule una estimación de tenacidad a la fractura promedio verdadera y también una predicción de tenacidad a la fractura tanto para un ángulo de 18 grados como para un ángulo de 22 grados y hágalo de modo que dé información sobre confiabilidad y precisión y luego interprete y compare las estimaciones y predicciones.

77. El artículo "Photocharge Effects in Dye Sensitized Ag[Br,I] Emulsions at Millisecond Range Exposures" (*Photographic Sci. and Engr.*, 1981: 138–144) da los datos adjuntos sobre x = % de absorción de luz a 5800 A y y = fotovoltaje máximo.

x	4.0	8.7	12.7	19.1	21.4
y	.12	.28	.55	.68	.85
x	24.6	28.9	29.8	30.5	
y	1.02	1.15	1.34	1.29	

a. Construya un diagrama de dispersión de estos datos. ¿Qué sugieren?

b. Suponiendo que el modelo de regresión lineal simple es apropiado, obtenga la ecuación de la recta de regresión estimada.

c. ¿Qué proporción de la variación observada del fotovoltaje máximo puede ser explicada por la relación de modelo?

d. Pronostique el fotovoltaje máximo cuando el % de absorción es de 19.1 y calcule el valor del residuo correspondiente.

e. Los autores del artículo manifiestan que existe una relación lineal útil entre el % de absorción y el fotovoltaje máximo. ¿Está de acuerdo? Realice una prueba formal.

f. Dé una estimación del cambio del fotovoltaje máximo esperado asociado con un incremento de 1% de la absorción de luz. Su estimación deberá informar sobre la precisión de la estimación.

g. Repita el inciso (f) del valor esperado del fotovoltaje máximo cuando el % de absorción de luz es 20.

78. En la sección 12.4, se presentó una fórmula para $V(\hat{\beta}_0 + \hat{\beta}_1 x^*)$ y un intervalo de confianza para $\beta_0 + \beta_1 x^*$. Considerando $x^* = 0$ se obtiene $\sigma^2_{\hat{\beta}_0}$ y un intervalo de confianza para β_0. Use los datos del ejemplo 12.11 para calcular la desviación estándar estimada de $\hat{\beta}_0$ y un intervalo de confianza de 95% para la intersección y de la línea de regresión verdadera.

79. Demuestre que $\text{SSE} = S_{yy} - \hat{\beta}_1 S_{xy}$, la cual da una fórmula computacional alternativa para SSE.

80. Suponga que x y y son variables positivas y que una muestra de n pares da $r \approx 1$. Si el coeficiente de correlación muestral se calcula para los pares (x, y^2), ¿será el valor resultante también aproximadamente 1? Explique.

81. Sean s_x y s_y las desviaciones estándar muestrales de las x y y observadas, respectivamente [así que $s_x^2 = \sum(x_i - \bar{x})^2/(n-1)$ y asimismo para s_y^2].

a. Demuestre que una expresión alternativa para la línea de regresión estimada $y = \hat{\beta}_0 + \hat{\beta}_1 x$ es

$$y = \bar{y} + r \cdot \frac{s_y}{s_x}(x - \bar{x})$$

b. Esta expresión para la línea de regresión puede ser interpretada como sigue. Suponga $r = .5$. ¿Cuál es entonces la y pronosticada con una x situada a 1 desviación estándar (s_x unidades) sobre la media de las x_i? Si r fuera 1, la predicción sería para que y quede a 1 desviación estándar sobre su media $\bar{y}$ pero como $r = .5$, se pronostica una y que está a sólo .5 desviación estándar ($.5s_y$ unidad) sobre $\bar{y}$. Con los datos del ejercicio 64 para un paciente cuya edad está a 1 desviación estándar por debajo de la edad promedio en la muestra, ¿a cuántas desviaciones estándar se pronostica que esté el ΔCBG pronosticado del paciente por encima o por debajo del ΔCBG promedio para la muestra?

82. Verifique que el estadístico t para probar H_0: $\beta_1 = 0$ en la sección 12.3 es idéntico al estadístico t en la sección 12.5 para probar H_0: $\rho = 0$.

83. Use la fórmula para calcular SSE para comprobar que $r^2 = 1 - \text{SSE/SST}$.

84. En la biofiltración de aguas residuales, se hace que el aire descargado por una planta de tratamiento pase a través de una membrana porosa húmeda que disuelve los contaminantes en el agua y los transforma en productos inocuos. Los datos adjuntos sobre x = temperatura de entrada (°C) y y = eficiencia de eliminación (%) fueron la base para una gráfica de dispersión que apareció en el artículo "Treatment of Mixed Hydrogen Sulfide and Organic Vapors in a Rock Medium Biofilter" (*Water Environment Research*, 2001: 426–435).

Obs	Temp	% de eliminación	Obs	Temp	% de eliminación
1	7.68	98.09	17	8.55	98.27
2	6.51	98.25	18	7.57	98.00
3	6.43	97.82	19	6.94	98.09
4	5.48	97.82	20	8.32	98.25
5	6.57	97.82	21	10.50	98.41
6	10.22	97.93	22	16.02	98.51
7	15.69	98.38	23	17.83	98.71
8	16.77	98.89	24	17.03	98.79
9	17.13	98.96	25	16.18	98.87
10	17.63	98.90	26	16.26	98.76
11	16.72	98.68	27	14.44	98.58
12	15.45	98.69	28	12.78	98.73
13	12.06	98.51	29	12.25	98.45
14	11.44	98.09	30	11.69	98.37
15	10.17	98.25	31	11.34	98.36
16	9.64	98.36	32	10.97	98.45

Las cantidades calculadas son $\sum x_i = 384.26$, $\sum y_i = 3149.04$, $\sum x_i^2 = 5099.2412$, $\sum x_i y_i = 37{,}850.7762$ y $\sum y_i^2 = 309{,}892.6548$.

a. Sugieren una gráfica de dispersión de los datos la pertinencia del modelo de regresión lineal de la muestra?
b. Ajuste el modelo de regresión lineal simple, obtenga una predicción puntual de la eficiencia de eliminación cuando la temperatura = 10.50 y calcule el valor del residuo correspondiente.
c. Aproximadamente ¿cuál es el tamaño de una desviación típica de puntos en la gráfica de dispersión con respecto a la recta de mínimos cuadrados?
d. ¿Qué proporción de la variación observada de la eficiencia de eliminación puede ser atribuida a la relación de modelo?
e. Estime el coeficiente de pendiente de modo que informe sobre confiabilidad y precisión e interprete su estimación.
f. Comunicación personal con los autores del artículo reveló que hubo una observación adicional que no estuvo incluida en su gráfica de dispersión: (6.53, 96.55). ¿Qué impacto tiene esta observación adicional en la ecuación de la recta de mínimos cuadrados y los valores de s y r^2?

85. Los procesos normales de incubación en acuacultura inevitablemente producen tensión en los peces, el que puede impactar negativamente el crecimiento, reproducción y calidad de la carne y susceptibilidad a enfermedades. Tal tensión se pone de manifiesto en los elevados y sostenidos niveles de corticosteroides. El artículo "Evaluation of Simple Instruments for the Measurement of Blood Glucose and Lactate and Plasma Protein as Stress Indicators en Fish" (*J. of the World Aquaculture Society*, 1999: 276−284) describió un experimento en el cual los peces se sometieron a un protocolo de tensión y luego se suspendió y se sometieron a prueba en varias ocasiones después de que se aplicó el protocolo. Los datos adjuntos sobre x = tiempo (min) y y = nivel de glucosa en sangre (mmol/L) se leyeron en la gráfica.

x	2	2	5	7	12	13	17	18	23	24	26	28
y	4.0	3.6	3.7	4.0	3.8	4.0	5.1	3.9	4.4	4.3	4.3	4.4
x	29	30	34	36	40	41	44	56	56	57	60	60
y	5.8	4.3	5.5	5.6	5.1	5.7	6.1	5.1	5.9	6.8	4.9	5.7

Use los métodos desarrollados en este capítulo para analizar los datos y escriba un breve reporte que resuma sus conclusiones (suponga que los investigadores están particularmente interesados en el nivel de glucosa 30 min después de la tensión).

86. El artículo "Evaluating the BOD POD for Assessing Body Fat in Collegiate Football Players" (*Medicine and Science in Sports and Exercise*, 1999: 1350–1356) reporta sobe un nuevo dispositivo de desplazamiento de aire para medir la grasa corporal. El procedimiento acostumbrado utiliza dispositivo de pesar hidrostático, el cual mide el porcentaje de masa corporal por medio del desplazamiento de agua. Los siguientes son datos representativos tomados de una gráfica que aparece en el artículo.

BOD	2.5	4.0	4.1	6.2	7.1	7.0	8.3	9.2	9.3	12.0	12.2
HW	8.0	6.2	9.2	6.4	8.6	12.2	7.2	12.0	14.9	12.1	15.3
BOD	12.6	14.2	14.4	15.1	15.2	16.3	17.1	17.9	17.9		
HW	14.8	14.3	16.3	17.9	19.5	17.5	14.3	18.3	16.2		

a. Use varios métodos para decidir si es plausible que las dos técnicas midan o promedien la misma cantidad de grasa.
b. Use los datos para desarrollar una forma de predecir un peso hidrostático a partir de una medición BOD POD e investigue la efectividad de tales predicciones.

87. Reconsidere la situación del ejercicio 73, en el cual x = contenido de austenita retenido utilizando un abrasivo de granate y y = pérdida por desgaste abrasivo se relacionaron vía el modelo de regresión lineal simple $Y = \beta_0 + \beta_1 x + \varepsilon$. Suponga que con un segundo tipo de abrasivo, estas variables también están relacionadas vía el modelo de regresión lineal simple $Y = \gamma_0 + \gamma_1 x + \varepsilon$ y que $V(\varepsilon) = \sigma^2$ para ambos tipos de abrasivo. Si el conjunto de datos se compone de n_1 observaciones del primer abrasivo y n_2 del segundo y si SSE_1 y SSE_2 denotan las dos sumas de cuadrados debido al error, entonces una estimación agrupada de σ^2 es $\hat{\sigma}^2 = (SSE_1 + SSE_2)/(n_1 + n_2 - 4)$. Sean SS_{x1} y SS_{x2} denote $\sum(x_i - \bar{x})^2$ con los datos del primero y segundo abrasivos, respectivamente. Una prueba de H_0: $\beta_1 - \gamma_1 = 0$ (pendientes iguales) está basada en el estadístico

$$T = \frac{\hat{\beta}_1 - \hat{\gamma}_1}{\hat{\sigma}\sqrt{\dfrac{1}{SS_{x1}} + \dfrac{1}{SS_{x2}}}}$$

Cuando H_0 es verdadera, T tiene una distribución t con $n_1 + n_2 - 4$ gl. Acepte las 15 observaciones utilizando los abrasivos alternativos con $SS_{x2} = 7152.5578$, $\hat{\gamma}_1 = .006845$, y $SSE_2 = .51350$. Empleando esto junto con los datos del ejercicio 73, realice una prueba de nivel .05 para ver si el cambio esperando en la pérdida por desgaste asociado con un 1% de aumento en el contenido de austenita es idéntico para los dos tipos de abrasivos.

Bibliografía

Draper, Norman y Harry Smith, *Applied Regression Analysis* (3a. ed.), Wiley, Nueva York, 1999. El libro más completo y autorizado sobre análisis de regresión actualmente en proceso de impresión.

Neter, John, Michael Kutner, Christopher Nachstsheim y William Wasserman, *Applied Linear Statistical Models* (5a. ed.), Irwin, Homewood, IL., 2005. Los primeros 14 capítulos constituyen un estudio extremadamente informativo y fácil de leer acerca del análisis de regresión.

13 Regresión múltiple y no lineal

INTRODUCCIÓN

El modelo probabilístico estudiado en el capítulo 12 especificó que el valor observado de la variable dependiente Y se desviaba de la función de regresión lineal $\mu_{Y \cdot x} = \beta_0 + \beta_1 x$ en una cantidad aleatoria. Aquí se consideran dos formas de generalizar el modelo de regresión lineal simple. La primera es sustituir $\beta_0 + \beta_1 x$ con una función no lineal de x, y la segunda es usar una función de regresión que comprenda más de una sola variable independiente. Después de ajustar una función de regresión de la forma seleccionada a la información dada, por supuesto que es importante tener métodos disponibles para hacer inferencias acerca de los parámetros del modelo seleccionado. No obstante, antes de usar estos métodos el analista de datos debe evaluar primero la validez del modelo seleccionado. En la sección 13.1 se estudian estos métodos, con base principalmente en un análisis gráfico de los residuos (las y observadas menos las pronosticadas), para verificar lo apropiado del modelo.

En la sección 13.2 se consideran funciones de regresión no lineales de una sola variable independiente x que son "intrínsecamente lineales". Con esto se quiere decir que es posible transformar una o las dos variables para que la relación entre las nuevas variables sea lineal. Se obtiene una clase alternativa de relaciones no lineales con el uso de funciones de regresión polinomiales de la forma $\mu_{Y \cdot x} = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_k x^k$; estos modelos polinomiales son el tema de la sección 13.3. El análisis de regresión múltiple comprende la construcción de modelos para relacionar y con dos o más variables independientes. El interés principal de la sección 13.4 está en la interpretación de varios modelos de regresión múltiple y en entender y usar la salida de regresión de varios paquetes estadísticos de computadoras. La última sección del capítulo examina algunas extensiones y dificultades al hacer modelos de regresión múltiple.

13.1 Aptitud y verificación del modelo

Una gráfica de los pares observados (x_i, y_i) es un primer paso necesario para decidir la forma de una relación matemática entre x y y. Es posible adaptar numerosas funciones que no sean una lineal ($y = b_0 + b_1x$) a los datos, usando ya sea el principio de mínimos cuadrados u otro método apropiado. Una vez que una función de la forma seleccionada se haya ajustado, es importante verificar el ajuste del modelo para ver si en verdad es apropiado. Una forma de estudiar el ajuste es sobreponer una gráfica de la función de mejor ajuste sobre el diagrama de dispersión de los datos. No obstante, cualquier inclinación o curvatura de la función de mejor ajuste puede ocultar algunos aspectos del ajuste que deben investigarse. Además, la escala en el eje vertical puede hacer difícil evaluar el grado al que los valores observados se desvían de las funciones de mejor ajuste.

Residuos y residuos estandarizados

Un método más eficaz de evaluar la exactitud del modelo es calcular los valores ajustados o pronosticados $\hat{y}_i$ y los residuos $e_i = y_i - \hat{y}_i$ y luego graficar varias funciones de estas cantidades calculadas. A continuación se examinan las gráficas para confirmar la selección de modelo o para indicaciones de que el modelo no es apropiado. Suponga que el modelo de regresión lineal simple es correcto, y sea $y = \hat{\beta}_0 + \hat{\beta}_1 x$ la ecuación de la recta de regresión estimada. Entonces el i-ésimo residuo es $e_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$. Para deducir propiedades de los residuos, se representa con $e_i = Y_i - \hat{Y}_i$ el i-ésimo residuo como una variable aleatoria (va) antes que en realidad se hagan observaciones. Entonces

$$E(Y_i - \hat{Y}_i) = E(Y_i) - E(\hat{\beta}_0 + \hat{\beta}_1 x_i) = \beta_0 + \beta_1 x_i - (\beta_0 + \beta_1 x_i) = 0 \quad (13.1)$$

Debido a que $\hat{Y}_i$ $(= \hat{\beta}_0 + \hat{\beta}_1 x_i)$ es una función lineal de las Y_j, así lo es $Y_i - \hat{Y}_i$ (los coeficientes dependen de las x_j). Así, la normalidad de las Y_j implica que cada residuo está normalmente distribuido. También se puede demostrar que

$$V(Y_i - \hat{Y}_i) = \sigma^2 \cdot \left[1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}}\right] \quad (13.2)$$

Si se sustituye σ^2 con s^2 y se toma la raíz cuadrada de la ecuación (13.2) resulta la desviación estándar estimada de un residuo.

Se estandariza ahora cada residuo al restar el valor medio (cero) y luego dividir entre la desviación estándar estimada.

Los **residuos estandarizados** están dados por

$$e_i^* = \frac{y_i - \hat{y}_i}{s\sqrt{1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}}}} \quad i = 1, \ldots, n \quad (13.3)$$

Si, por ejemplo, un residuo estandarizado particular es 1.5, entonces el residuo en sí es 1.5 desviaciones estándar (estimadas) mayor de lo que se esperaría por ajustar el modelo correcto. Nótese que las varianzas de los residuos difieren entre sí. De hecho, ya que hay un signo $-$ delante de $(x_i - \bar{x})^2$, la varianza de un residuo disminuye a medida que x_i está más lejos del centro de los datos $\bar{x}$. Intuitivamente, esto se debe a que la recta de mínimos cuadrados se jala hacia una observación cuyo valor x_i se encuentra más a la derecha o a la izquierda de otras observaciones en la muestra. El cálculo de las e_i^* puede ser tedioso, pero

los paquetes computarizados de estadísticas de más uso dan estos valores de manera automática y pueden construir varias gráficas donde esos valores están comprendidos.

Ejemplo 13.1 El ejercicio 19 del capítulo 12 presentó datos acerca de x = tasa de liberación debido a área del quemador y y = tasa de emisión de NO_x. Aquí se reproducen los datos y los valores ajustados, residuos y residuos estandarizados. La recta de regresión estimada es $y = -45.55 + 1.71x$, y $r^2 = .961$. Observe que los residuos estandarizados no son un múltiplo constante de los residuos debido a que las varianzas residuales difieren unas de otras.

x_i	y_i	$\hat{y}_i$	e_i	e_i^*
100	150	125.6	24.4	.75
125	140	168.4	−28.4	−.84
125	180	168.4	11.6	.35
150	210	211.1	−1.1	−.03
150	190	211.1	−21.1	−.62
200	320	296.7	23.3	.66
200	280	296.7	−16.7	−.47
250	400	382.3	17.7	.50
250	430	382.3	47.7	1.35
300	440	467.9	−27.9	−.80
300	390	467.9	−77.9	−2.24
350	600	553.4	46.6	1.39
400	610	639.0	−29.0	−.92
400	670	639.0	31.0	.99

■

Gráficas de diagnóstico

Las gráficas básicas que numerosos expertos en estadística recomiendan para una evaluación de la validez y utilidad de un modelo son las siguientes:

1. e_i^* (o e_i) sobre el eje vertical contra x_i en el eje horizontal
2. e_i^* (o e_i) sobre el eje vertical contra $\hat{y}_i$ en el eje horizontal
3. $\hat{y}_i$ sobre el eje vertical contra y_i en el eje horizontal
4. Una gráfica de probabilidad normal de los residuos estandarizados

Las gráficas 1 y 2 se denominan **gráficas de residuos** (contra la variable independiente y valores ajustados, respectivamente), en tanto que la gráfica 3 está ajustada contra valores observados.

Si la gráfica 3 da puntos cercanos a la recta de 45° [pendiente +1 que pasa por (0, 0)], entonces la función de regresión estimada da predicciones precisas de los valores que se observan en realidad. Así, la gráfica 3 proporciona una evaluación visual de la efectividad del modelo para hacer predicciones. Siempre que el modelo sea correcto, ninguna gráfica de residuos debe exhibir formas distintas. Los residuos deben estar distribuidos al azar alrededor de 0 según una distribución normal, de manera que con excepción de unos cuantos, todos los residuos estandarizados deben encontrarse entre −2 y +2 (es decir, todos excepto unos cuantos a no más de dos desviaciones estándar de su valor esperado 0). La gráfica de residuos estandarizados contra $\hat{y}$ es en realidad una combinación de las otras dos gráficas, mostrando implícitamente la forma en que varían los residuos con x y cómo se comparan los valores ajustados con valores observados. Esta última gráfica es la que se recomienda con más frecuencia para análisis de regresión múltiple. La gráfica 4 permite al analista evaluar la factibilidad de la suposición de que ϵ tiene una distribución normal.

Ejemplo 13.2
(Continuación del ejemplo 13.1)

La figura 13.1 presenta un diagrama de dispersión de los datos y las cuatro gráficas recomendadas. La gráfica de $\hat{y}$ contra y confirma la impresión dada por r^2 de que x es eficaz en la predicción de y y también indica que no hay y observada para la que el valor predicho esté muy lejos de la marca. Ambas gráficas de residuos no muestran una figura poco común ni valores discrepantes. Hay un residuo estandarizado que está ligeramente fuera del intervalo $(-2, 2)$, pero esto no es sorprendente en una muestra de tamaño 14. La gráfica de probabilidad normal de los residuos estandarizados es razonablemente recta. En resumen, las gráficas no dejan remordimiento acerca de lo apropiado de una relación lineal sencilla o el ajuste a la información dada.

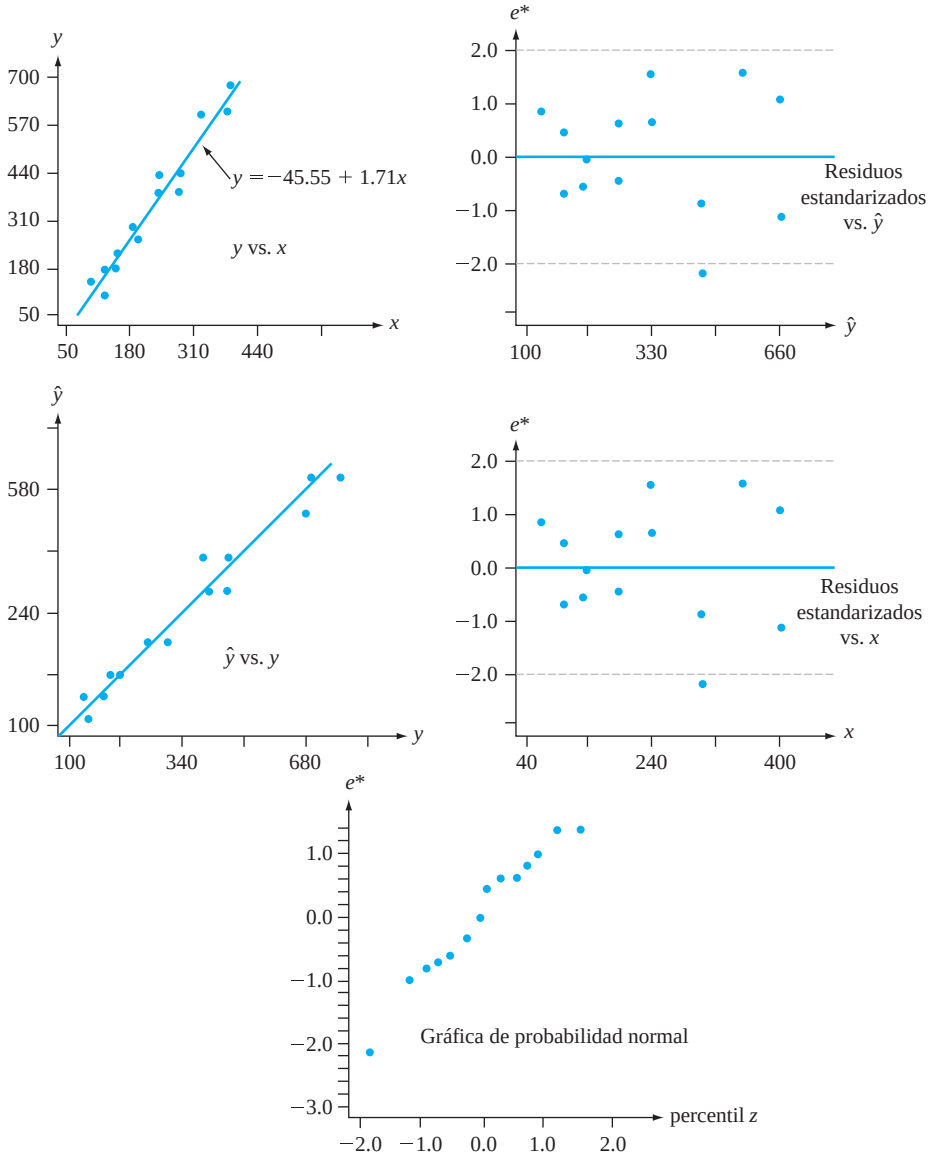


Figura 13.1 Gráficas para los datos del ejemplo 13.1 ■

Dificultades y soluciones

Aun cuando se espera que nuestro análisis dé gráficas como las de la figura 13.1, con gran frecuencia dichas gráficas sugerirán una o más de las siguientes dificultades:

1. Una relación probabilística no lineal entre x y y es apropiada.
2. La varianza de ϵ (y de Y) no es una σ^2 constante sino que depende de x.
3. El modelo seleccionado se ajusta bien a los datos, excepto para unos pocos valores discrepantes de datos o resultados aislados, que pueden haber tenido gran influencia en la selección de la función de mejor ajuste.
4. El término de error ϵ no tiene una distribución normal.
5. Cuando el subíndice i indica el orden temporal de las observaciones en tiempo, las ϵ_i exhiben dependencia en el tiempo.
6. Una o más variables independientes relevantes se han omitido del modelo.

La figura 13.2 presenta gráficas de residuos correspondientes a los elementos 1–3, 5 y 6. En el capítulo 4, se estudiaron figuras en gráficas de probabilidad normales que arrojan duda sobre la suposición de una distribución normal subyacente. Nótese que los residuos de los datos de la figura 13.2(d), con el punto circulado incluido, por sí mismos no sugerirían un análisis ulterior, pero cuando se ajusta una nueva recta con ese punto borrado, la nueva recta difiere considerablemente de la recta original. Este tipo de conducta es más difícil de identificar en regresión múltiple. Lo más probable es que surja cuando haya un solo (o muy pocos) punto(s) de dato con valor(es) variable(s) independiente(s) muy alejados del resto de los datos.

A continuación se indica brevemente de qué soluciones se dispone para los tipos de dificultades. Para un análisis más completo debe consultarse una o más de las referencias sobre análisis de regresión. Si la gráfica de residuos se ve como la de la figura 13.2(a), que exhibe una figura curva, entonces puede ajustarse una función no lineal de x.

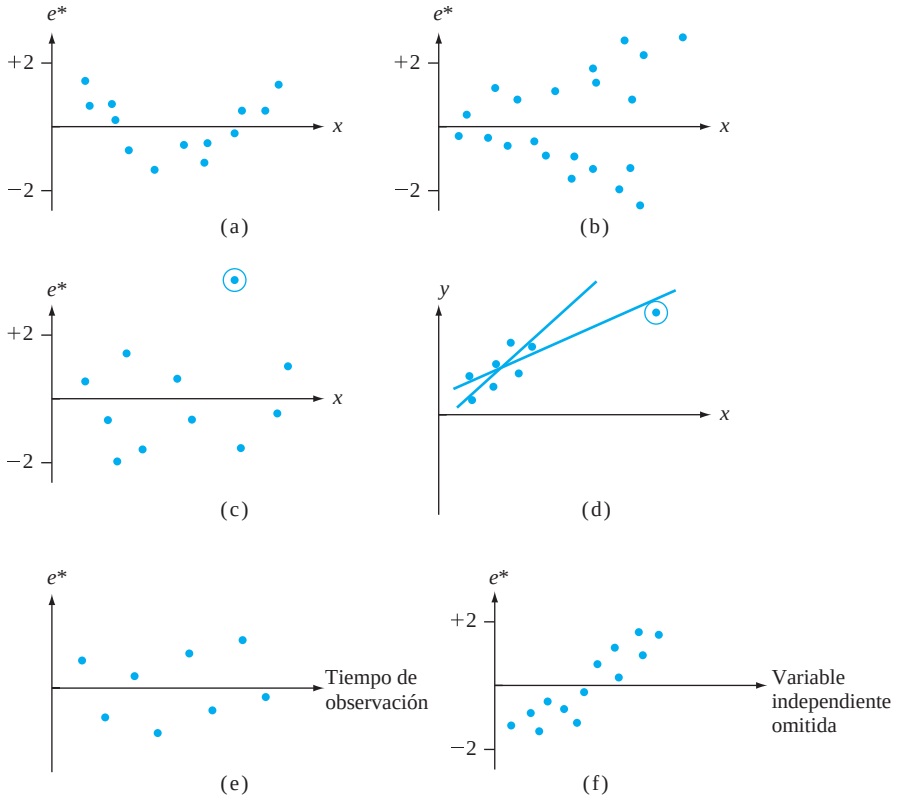


Figura 13.2 Gráficas que indican anormalidad en datos: (a) relación no lineal; (b) varianza no constante; (c) observación discrepante; (d) observación con gran influencia; (e) dependencia de errores; (f) variable omitida

La gráfica de residuos de la figura 13.2(b) sugiere que, aun cuando puede ser razonable una relación de línea recta, la suposición de que $V(Y_i) = \sigma^2$ para cada i es de dudosa validez. Cuando las suposiciones del capítulo 12 son válidas, se puede demostrar que entre todos los estimadores no sesgados de β_0 y β_1, los estimadores de mínimos cuadrados ordinarios tienen varianza mínima. Estos estimadores dan igual valor a cada (x_i, Y_i). Si la varianza de Y aumenta con x, entonces las Y_i para x_i grandes deben tener menos valor que aquellas con x_i pequeñas. Esto sugiere que β_0 y β_1 deben estimarse al minimizar

$$f_w(b_0, b_1) = \sum w_i[y_i - (b_0 + b_1 x_i)]^2 \qquad (13.4)$$

donde las w_i son valores que decrecen con x_i creciente. La reducción al mínimo de la expresión (13.4) da estimaciones de **mínimos cuadrados ponderados.** Por ejemplo, si la desviación estándar de Y es proporcional a x (para $x > 0$), es decir, $V(Y) = kx^2$, entonces se puede demostrar que los valores $w_i = 1/x_i^2$ dan mejores estimadores de β_0 y β_1. Los libros de John Neter y otros y de S. Chatterjee y Bertram Price contienen más detalle (vea la bibliografía del capítulo). Los mínimos cuadrados ponderados los emplean con frecuencia expertos en econometría (economistas que usan métodos estadísticos) para estimar parámetros.

Cuando las gráficas u otra evidencia sugieren que el conjunto de datos contiene resultados aislados o puntos que tienen gran influencia en el ajuste resultante, un posible método es omitir estos puntos aislados y recalcular la ecuación de regresión estimada. Es seguro que esto sería correcto si se encontrara que los resultados aislados aparecieron por errores al registrar valores de datos o de errores experimentales. Si no se puede hallar una causa para los resultados aislados, es deseable informar la ecuación estimada con y sin haber omitido los resultados aislados. Otro método adicional es retener posibles resultados aislados pero sólo para usar un principio de estimación que pone relativamente menos peso en valores aislados del que da el principio de mínimos cuadrados. Uno de estos principios es el MAD (minimizar desviaciones absolutas), que selecciona $\hat{\beta}_0$ y $\hat{\beta}_1$ para minimizar $\sum |y_i - (b_0 + b_1 x_i)|$. A diferencia de las estimaciones de mínimos cuadrados, no hay fórmulas exactas para las estimaciones MAD; sus valores deben hallarse con el uso de procedimientos computacionales iterativos. Estos procedimientos también se usan cuando se sospecha que las ϵ_i tienen una distribución que no es normal y que, en cambio, tiene "colas pesadas" (lo cual hace más probable para la distribución normal que valores discrepantes entren en la muestra); los procedimientos de regresión robustos son aquellos que producen estimaciones confiables para una amplia variedad de distribuciones de error subyacentes. Los estimadores de mínimos cuadrados no son robustos en la misma forma que la media muestral $\overline{X}$ no es un estimador robusto para μ.

Cuando una gráfica sugiere dependencia del tiempo en los términos de error, un análisis apropiado puede comprender una transformación de las y o un modelo que en forma explícita incluya una variable de tiempo. Por último, una gráfica como la de la figura 13.2(f), que presenta un patrón en los residuos cuando se traza contra una variable omitida, sugiere que debe considerarse un modelo de regresión múltiple que incluya la variable previamente omitida.

EJERCICIOS Sección 13.1 (1–14)

1. Suponga que las variables x = distancia de viaje al trabajo y y = tiempo de viaje al trabajo están relacionadas de acuerdo con el modelo de regresión lineal simple con $\sigma = 10$.

a. Si se hacen $n = 5$ observaciones en los valores x de $x_1 = 5$, $x_2 = 10$, $x_2 = 15$, $x_2 = 20$, y $x_5 = 25$, calcule las desviaciones estándar de los cinco residuos correspondientes.

b. Repita el inciso (a) para $x_1 = 5$, $x_2 = 10$, $x_3 = 15$, $x_4 = 20$, y $x_5 = 50$.

c. ¿Qué implican los resultados de los incisos (a) y (b) acerca de la desviación de la recta estimada a partir de la observación hecha en el valor x máximo muestreado?

2. Los valores x y residuos estandarizados para los datos de flujo de cloro/(velocidad de grabado), del ejercicio 52 (sección 12.4), se muestran en la tabla siguiente. Construya una gráfica de residuos estandarizada y comente sobre su aspecto.

x	1.50	1.50	2.00	2.50	2.50
e^*	.31	1.02	−1.15	−1.23	.23
x	3.00	3.50	3.50	4.00	
e^*	.73	−1.36	1.53	.07	

3. El ejemplo 12.6 presentó los residuos de una regresión lineal simple de contenido de humedad y sobre la rapidez de filtración x.

a. Trace los residuos en función de x. ¿La gráfica resultante sugiere que una función de regresión de línea recta es una opción razonable de modelo? Explique su razonamiento.

b. Usando $s = .665$, calcule los valores de los residuos estandarizados. ¿Es $e_i^* \approx e_i$ para $i = 1, \ldots, n$, o no están las e_i^* cerca de ser proporcionales a las e_i?

c. Trace los residuos estandarizados en función de x. ¿Difiere esta gráfica significativamente en su aspecto general con respecto a la gráfica del inciso (a)?

4. La resistencia al desgaste de ciertos componentes de reactores nucleares hechos de Zircaloy-2 se determina en parte por las propiedades de la capa de óxido. La siguiente información aparece en un artículo que propuso un nuevo método de prueba no destructivo para vigilar el grosor de la capa ("Monitoring of Oxide Layer Thickness on Zircaloy-2 by the Eddy Current Test Method", *J. of Testing and Eval.*, 1987: 333–336). Las variables son x = grosor de la capa de óxido (μm) y y = respuesta de la corriente parásita o turbulenta (unidades arbitrarias).

x	0	7	17	114	133
y	20.3	19.8	19.5	15.9	15.1
x	142	190	218	237	285
y	14.7	11.9	11.5	8.3	6.6

a. Los autores resumieron la relación al dar la ecuación de la recta de mínimos cuadrados como $y = 20.6 - .047x$. Calcule y trace los residuos en función de x y luego comente sobre lo apropiado del modelo de regresión lineal simple.

b. Use $s = .7921$ para calcular los residuos estandarizados de una regresión lineal simple. Construya una gráfica de residuos estandarizada y comente. También construya una gráfica de probabilidad normal y comente.

5. Cuando desciende la temperatura del aire, el agua de un río se hace muy fría y se forman cristales de hielo. Este hielo puede afectar de manera significativa la hidráulica de un río. El artículo "Laboratory Study of Anchor Ice Growth" (*J. of Cold Regions Engr.*, 2001: 60–66) describió un experimento en el que el grosor del hielo (mm) se estudió como función del tiempo transcurrido (h) bajo condiciones especificadas. La información siguiente se leyó de una gráfica del artículo: $n = 33$; $x = .17, .33, .50, .67, \ldots, 5.50$; $y = .50, 1.25, 1.50, 2.75, 3.50, 4.75, 5.75, 5.60, 7.00, 8.00, 8.25, 9.50, 10.50, 11.00, 10.75, 12.50, 12.25, 13.25, 15.50, 15.00, 15.25, 16.25, 17.25, 18.00, 18.25, 18.15, 20.25, 19.50, 20.00, 20.50, 20.60, 20.50, 19.80$.

a. El valor r^2 resultante de un ajuste de mínimos cuadrados es .977. Interprete este valor y comente sobre lo apropiado de suponer una relación lineal aproximada.

b. Los residuos, escritos en el mismo orden que los valores x, son:

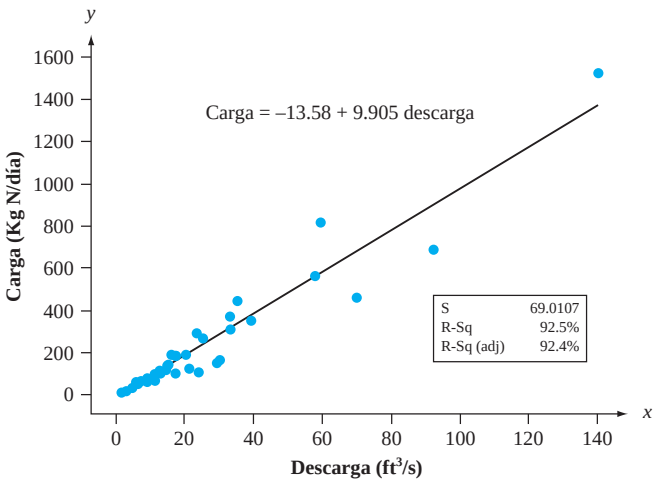
−1.03	−0.92	−1.35	−0.78	−0.68	−0.11	0.21
−0.59	0.13	0.45	0.06	0.62	0.94	0.80
−0.14	0.93	0.04	0.36	1.92	0.78	0.35
0.67	1.02	1.09	0.66	−0.09	1.33	−0.10
−0.24	−0.43	−1.01	−1.75	−3.14		

Grafique los residuos contra el tiempo transcurrido. ¿Qué sugiere la gráfica?

6. El gráfico de dispersión anexo se basa en datos proporcionados por los autores del artículo "Spurious Correlation in the USEPA Rating Curve Method for Estimating Pollutant Loads" (*J. of Envir. Engr.*, 2008: 610–618), aquí la descarga es en ft^3/s en lugar de m^3/s utilizados en el artículo. El punto en el extremo derecho de la gráfica corresponde a la observación (140, 1529.35). El resultado residual estandarizado es de 3.10. Minitab marca la observación con una R para grandes residuales y una X para la observación potencialmente influyente. Aquí hay alguna información sobre la pendiente estimada:

	Muestra completa	**(140, 1529.35) eliminada**
$\hat{\beta}_1$	9.9050	8.8241
$s_{\hat{\beta}_1}$	.3806	.4734

¿Esta observación parece haber tenido un impacto sustancial en la pendiente estimada? Explique.



7. Emparedados o "sándwiches" de paneles de compuesto son ampliamente utilizados en diversas aplicaciones estructurales aeroespaciales, tales como las costillas, aletas y timones. El artículo "Core Crush Problem in Manufacturing of Composite Sandwich Structures: Mechanisms and Solutions" (*Amer. Inst. of Aeronautics and Astronautics J.*, 2006: 901–907), ajusta a una recta los datos siguientes en x = espesor preimpregnado = (mm) y = núcleo aplastado (%):

x	.246	.250	.251	.251	.254	.262	.264	.270
y	16.0	11.0	15.0	10.5	13.5	7.5	6.1	1.7
x	.272	.277	.281	.289	.290	.292	.293	
y	3.6	0.7	0.9	1.0	0.7	3.0	3.1	

a. Ajuste el modelo de regresión lineal simple. ¿Qué proporción de la variación observada en el núcleo aplastado puede ser atribuida a la relación del modelo?

b. Construya un diagrama de dispersión. ¿El argumento sugiere que una relación probabilística lineal es la adecuada?

c. Obtenga los residuos y los residuos estandarizados y luego construya los gráficos de residuos. ¿Qué sugieren estas gráficas? ¿Qué tipo de función debe proporcionar un mejor ajuste a los datos que el de una línea recta?

8. El registro continuo de la pulsación cardiaca se puede usar para obtener información acerca del nivel de intensidad de ejercicio o esfuerzo físico durante una participación deportiva, trabajo, u otras actividades diarias. El artículo "The Relationship Between Heart Rate and Oxygen Uptake During Non-Steady State Exercise" (*Ergonomics*, 2000: 1578–1592) publicó un estudio para investigar el uso de la respuesta del ritmo cardiaco (x, como porcentaje del ritmo máximo) para predecir la toma de oxígeno (y, como porcentaje de toma máxima) durante el ejercicio. La información siguiente es de una gráfica del artículo.

RC	43.5	44.0	44.0	44.5	44.0	45.0	48.0	49.0
VO_2	22.0	21.0	22.0	21.5	25.5	24.5	30.0	28.0
RC	49.5	51.0	54.5	57.5	57.7	61.0	63.0	72.0
VO_2	32.0	29.0	38.5	30.5	57.0	40.0	58.0	72.0

Use un paquete computacional de estadística para efectuar un análisis de regresión lineal simple, poniendo particular atención a la presencia de cualesquiera observaciones poco comunes o que influyan.

9. Considere los siguientes cuatro conjuntos de datos (x, y); los tres primeros tienen los mismos valores de x, de modo que estos valores aparecen sólo una vez. (Frank Anscombe, "Graphs in Statistical Analysis", *Amer. Statistician*, 1973: 17-21):

Conjunto de datos	**1–3**	**1**	**2**	**3**	**4**	**4**
Variable	x	y	y	y	x	y
	10.0	8.04	9.14	7.46	8.0	6.58
	8.0	6.95	8.14	6.77	8.0	5.76
	13.0	7.58	8.74	12.74	8.0	7.71
	9.0	8.81	8.77	7.11	8.0	8.84
	11.0	8.33	9.26	7.81	8.0	8.47
	14.0	9.96	8.10	8.84	8.0	7.04
	6.0	7.24	6.13	6.08	8.0	5.25
	4.0	4.26	3.10	5.39	19.0	12.50
	12.0	10.84	9.13	8.15	8.0	5.56
	7.0	4.82	7.26	6.42	8.0	7.91
	5.0	5.68	4.74	5.73	8.0	6.89

Para cada uno de estos cuatro conjuntos de datos, los valores de las estadísticas de resumen $\sum x_i$, $\sum x_i^2$, $\sum y_i$, $\sum y_i^2$ y $\sum x_i y_i$ son prácticamente idénticos, de modo que todas las cantidades calculadas de estos cinco serán idénticas en esencia para los cuatro conjuntos: la recta de mínimos cuadrados ($y = 3 + .5x$), SSE, s^2, r^2, intervalos t, estadísticos t, etcétera. Los resúmenes estadísticos no dan una forma de distinguir entre los cuatro conjuntos de datos. Con base en un diagrama de dispersión y una gráfica de residuos para cada conjunto, comente sobre lo apropiado o no de ajustar un modelo de línea recta; incluya en sus comentarios cualesquiera sugerencias específicas sobre cómo es que un "análisis de línea recta" podría modificarse o calificarse.

10. **a.** Demuestre que $\sum_{i=1}^{n} e_i = 0$ cuando las e_i son los residuos de una regresión lineal simple.

b. Los residuos de una regresión lineal simple ¿son independientes entre sí, están positivamente correlacionados, o negativamente correlacionados? Explique.

c. Demuestre que $\sum_{i=1}^{n} x_i e_i = 0$ para los residuos de una regresión lineal simple. (Este resultado junto con el inciso (a) muestra que hay dos restricciones lineales en las e_i, resultando en una pérdida de 2 grados de libertad cuando se usa el cuadrado de los residuos para estimar σ^2.)

d. ¿Es cierto que $\sum_{i=1}^{n} e_i^* = 0$? Dé una prueba o un ejemplo contrario.

11. **a.** Exprese el i-ésimo residuo $Y_i - \hat{Y}_i$, donde $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, en la forma $\sum c_j Y_j$, una función lineal de las Y_j. A continuación use reglas de varianza para verificar que $V(Y_i - \hat{Y}_i)$ está dada por la expresión (13.2).

b. Se puede demostrar que $\hat{Y}_i$ y $Y_i - \hat{Y}_i$ (el i-ésimo valor pronosticado y residuo) son independientes entre sí. Use este hecho, la relación $Y_i = \hat{Y}_i + (Y_i - \hat{Y}_i)$, y la expresión para $V(\hat{Y})$ de la sección 12.4 para otra vez verificar la expresión (13.2).

c. Cuando x_i se aleja de $\bar{x}$, ¿qué ocurre a $V(\hat{Y}_i)$ y a $V(Y_i - \hat{Y}_i)$?

12. **a.** ¿Podría una regresión lineal dar por resultado 23, −27, 5, 17, −8, 9 y 15? ¿Por qué sí o por qué no?

b. ¿Podría una regresión lineal resultar en residuos 23, −27, 5, 17, −8, −12 y 2 correspondientes a los valores x de 3, −4, 8, 12, −14, −20, y 25? ¿Por qué sí o por qué no? [*Sugerencia:* vea el ejercicio 10.]

13. Recuerde que $\hat{\beta}_0 + \hat{\beta}_1 x$ tiene una distribución normal con valor esperado $\beta_0 + \beta_1 x$ y varianza

$$\sigma^2 \left\{ \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right\}$$

de modo que

$$Z = \frac{\hat{\beta}_0 + \hat{\beta}_1 x - (\beta_0 + \beta_1 x)}{\sigma \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right)^{1/2}}$$

tiene una distribución normal estándar. Si $S = \sqrt{\text{SSE}/(n-2)}$ se sustituye por σ, la variable resultante tiene una distribución t con $n-2$ grados de libertad. Por analogía, ¿cuál es la distribución de cualquier residuo estandarizado en particular? Si $n = 25$, ¿cuál es la probabilidad de que un residuo estandarizado en particular caiga fuera del intervalo $(-2.50, 2.50)$?

14. Si hay al menos un valor de x en el que más de una observación se haya hecho, hay un procedimiento formal de prueba para probar H_0: $\mu_{Y \cdot x} = \beta_0 + \beta_1 x$ para algunos valores β_0, β_1 (la función de regresión verdadera es lineal)

contra

H_a: H_0 no es verdadera (la función de regresión verdadera no es lineal)

Suponga que se hacen observaciones en $x_1, x_2, \ldots, x_c$. Denote con $Y_{11}, Y_{12}, \ldots, Y_{1n_1}$ las n_1 observaciones cuando $x = x_1; \ldots; Y_{c1}, Y_{c2}, \ldots, Y_{cn_c}$ denota las n_c observaciones cuando $x = x_c$. Con $n = \sum n_i$ (el número total de observaciones), SSE tiene $n-2$ grados de libertad. Se descompone SSE en dos partes, SSPE (error puro) y SSLF (falta de ajuste) como sigue:

$$\text{SSPE} = \sum_i \sum_j (Y_{ij} - \bar{Y}_{i\cdot})^2$$
$$= \sum\sum Y_{ij}^2 - \sum n_i \bar{Y}_{i\cdot}^2$$
$$\text{SSLF} = \text{SSE} - \text{SSPE}$$

Las n_i observaciones en x_i contribuyen con $n_i - 1$ grados de libertad a SSPE, de modo que el número de grados de libertad para SSPE es $\Sigma_i(n_i - 1) = n - c$ y los grados de libertad para SSLF son $n - 2 - (n - c) = c - 2$. Sea $\text{MSPE} = \text{SSPE}/(n-c)$ y $\text{MSLF} = \text{SSLF}/(c-2)$. Entonces se puede demostrar que mientras $E(\text{MSPE}) = \sigma^2$ ya sea que H_0 sea o no verdadera, $E(\text{MSLF}) = \sigma^2$ si H_0 es verdadera y $E(\text{MSLF}) > \sigma^2$ si H_0 es falsa.

Estadístico de prueba: $F = \dfrac{\text{MSLF}}{\text{MSPE}}$

Región de rechazo: $f \geq F_{\alpha, c-2, n-c}$

Los datos siguientes provienen del artículo "Changes in Growth Hormone Status Related to Body Weight of Growing Cattle" (*Growth*, 1977: 241–247), con x = peso corporal y y = rapidez de eliminación metabólica/peso corporal.

x	110	110	110	230	230	230	360
y	235	198	173	174	149	124	115
x	360	360	360	505	505	505	505
y	130	102	95	122	112	98	96

(Así, $c = 4$, $n_1 = n_2 = 3$, $n_3 = n_4 = 4$.)

a. Pruebe H_0 contra H_a al nivel .05 usando la prueba de falta de ajuste que se acaba de describir.

b. Un diagrama de dispersión de los datos, ¿sugiere que la relación entre x y y es lineal? ¿Cómo se compara esto con el resultado del inciso (a)? (Una función de regresión no lineal se utilizó en el artículo.)

13.2 Regresión con variables transformadas

La necesidad de una alternativa para el modelo lineal $Y = \beta_0 + \beta_1 x + \epsilon$ puede ser sugerida ya sea por un argumento teórico o al examinar gráficas de diagnóstico desde un análisis de regresión lineal. En cualquiera de estos casos, es deseable escoger un modelo cuyos parámetros se puedan estimar con facilidad. Una clase importante de estos modelos se especifica por medio de funciones que sean "intrínsecamente lineales".

DEFINICIÓN

Una función que relacione y con x es **intrínsecamente lineal** si por medio de una transformación de x y/o y, la función se puede expresar como $y' = \beta_0 + \beta_1 x'$, donde $x' =$ la variable independiente transformada y $y' =$ la variable dependiente transformada.

En la tabla 13.1 se dan cuatro de las funciones intrínsecamente lineales más útiles. En cada caso, la transformación apropiada es una transformación logarítmica, ya sea de logaritmos de base 10 o naturales (base e), o una transformación recíproca. Unas gráficas representativas de las cuatro funciones aparecen en la figura 13.3.

Para una relación de función exponencial, sólo y se transforma para alcanzar linealidad, mientras que, para una relación de función de potencia, tanto x como y se transforman. Debido a que la variable x está en el exponente de una relación exponencial, y crece

Tabla 13.1 Funciones intrínsecamente lineales útiles*

Función	Transformación(es) para linealizar	Forma lineal
a. Exponencial: $y = \alpha e^{\beta x}$	$y' = \ln(y)$	$y' = \ln(\alpha) + \beta x$
b. Potencia: $y = \alpha x^{\beta}$	$y' = \log(y), x' = \log(x)$	$y' = \log(\alpha) + \beta x'$
c. $y = \alpha + \beta \cdot \log(x)$	$x' = \log(x)$	$y = \alpha + \beta x'$
d. Recíproca: $y = \alpha + \beta \cdot \frac{1}{x}$	$x' = \frac{1}{x}$	$y = \alpha + \beta x'$

**Cuando aparece log (·), se pueden usar logaritmos de base 10 o de base *e*.

(si $\beta > 0$) o decrece (si $\beta < 0$) en forma mucho más rápida cuando x aumenta más de lo que es el caso para la función de potencia, aun cuando en un breve intervalo de valores de x puede ser difícil distinguir entre las dos funciones. Ejemplos de funciones que no son intrínsecamente lineales son $y = \alpha + \gamma^{\beta x}$ y $y = \alpha + \gamma x^{\beta}$.

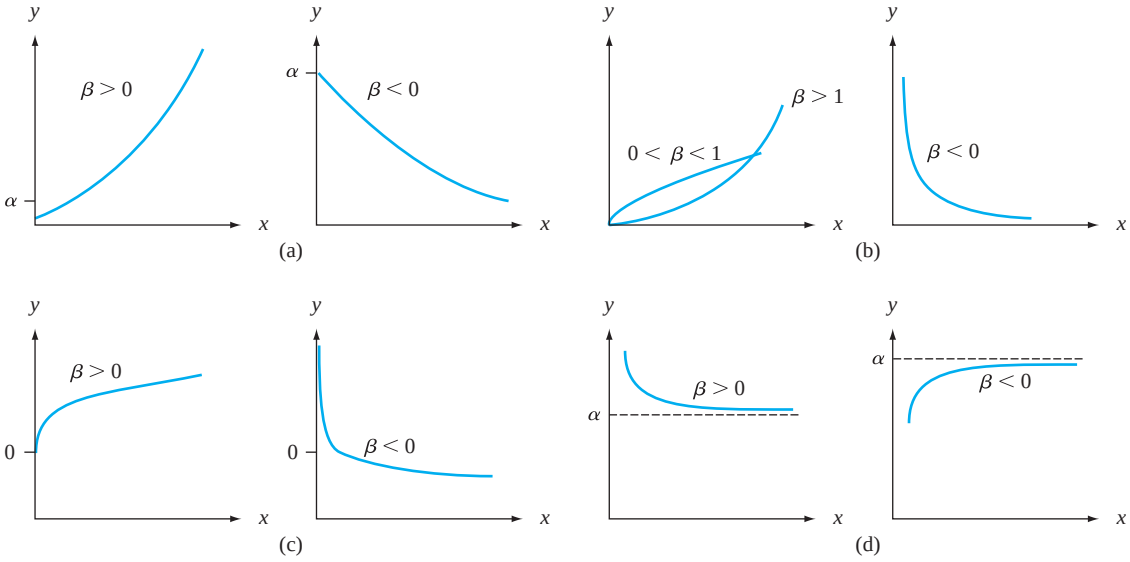


Figura 13.3 Gráficas de las funciones intrínsecamente lineales dadas en la tabla 13.1

Las funciones intrínsecamente lineales llevan de manera directa a modelos probabilísticos que, aun cuando no son lineales en x como función, tienen parámetros cuyos valores se estiman con facilidad usando mínimos cuadrados ordinarios.

DEFINICIÓN

Un modelo probabilístico que relaciona Y con x es **intrínsecamente lineal** si, por medio de una transformación en Y y/o x, se puede reducir a un modelo probabilístico lineal $Y' = \beta_0 + \beta_1 x' + \epsilon'$.

Los modelos probabilísticos intrínsecamente lineales que corresponden a las cuatro funciones de la tabla 13.1 son los siguientes:

a. $Y = \alpha e^{\beta x} \cdot \epsilon$, un modelo multiplicativo exponencial, de modo que $\ln(Y) = Y' = \beta_0 + \beta_1 x' + \epsilon'$ con $x' = x$, $\beta_0 = \ln(\alpha)$, $\beta_1 = \beta$, y $\epsilon' = \ln(\epsilon)$.

b. $Y = \alpha x^{\beta} \cdot \epsilon$, un modelo multiplicativo de potencia, de modo que $\log(Y) = Y' = \beta_0 + \beta_1 x' + \epsilon'$ con $x' = \log(x)$, $\beta_0 = \log(x) + \epsilon$ y $\epsilon' = \log(\epsilon)$.

c. $Y = \alpha + \beta \log(x) + \epsilon$, de modo que $x' = \log(x)$ linealiza de inmediato el modelo.

d. $Y = \alpha + \beta \cdot 1/x + \epsilon$, de modo que $x' = 1/x$ da un modelo lineal.

Los modelos aditivos exponencial y de potencia, $Y = \alpha e^{\beta x} + \epsilon$ y $Y = \alpha x^{\beta} + \epsilon$, no son intrínsecamente lineales. Nótese que (a) y (b) requieren una transformación en Y y, como resultado, una transformación en la variable de error ϵ. De hecho, si ϵ tiene una distribución lognormal (véase capítulo 4) con $E(\epsilon) = e^{\sigma^2/2}$ y $V(\epsilon) = \tau^2$ independientes de x, entonces los modelos transformados para (a) y (b) van a satisfacer todas las suposiciones del capítulo 12 con respecto al modelo probabilístico lineal; esto a su vez implica que todas las inferencias para los parámetros del modelo transformado con base en estas suposiciones será válido. Si σ^2 es pequeña, $\mu_{Y \cdot x} \approx \alpha e^{\beta x}$ en (a) o αx^{β} en (b).

La ventaja principal de un modelo intrínsecamente lineal es que los parámetros β_0 y β_1 del modelo transformado se pueden estimar de inmediato, con el uso del principio de mínimos cuadrados, con sólo sustituir x' y y' en las fórmulas estimadoras:

$$\hat{\beta}_1 = \frac{\sum x_i' y_i' - \sum x_i' \sum y_i'/n}{\sum (x_i')^2 - (\sum x_i')^2/n}$$

$$\hat{\beta}_0 = \frac{\sum y_i' - \hat{\beta}_1 \sum x_i'}{n} = \bar{y}' - \hat{\beta}_1 \bar{x}' \qquad (13.5)$$

Los parámetros del modelo original no lineal se pueden estimar entonces al transformar de nuevo $\hat{\beta}_0$ y/o $\hat{\beta}_1$ si es necesario. Una vez calculado el intervalo de predicción para y' cuando $x' = x'^*$, la inversión de la transformación da un intervalo de predicción (IP) para y misma. En los casos (a) y (b), cuando σ^2 es pequeña, un intervalo de confianza (IC) para $\mu_{Y \cdot x^*}$ resulta de tomar los antilog de los límites del IC por $\beta_0 + \beta_1 x'^*$ (estrictamente hablando, tomar antilogaritmos da un IC para la *mediana* de la distribución Y, es decir, para $\tilde{\mu}_{Y \cdot x^*}$. Debido a que la distribución lognormal está sesgada de manera positiva, $\mu > \tilde{\mu}$; las dos son aproximadamente iguales si σ^2 es cercana a 0.)

Ejemplo 13.3 La ecuación de Taylor para la duración y de herramientas como función del tiempo de corte x indica que $xy^c = k$ o bien, lo que es equivalente, que $y = \alpha x^{\beta}$. El artículo "The Effect of Experimental Error on the Determination of Optimum Metal Cutting Conditions" (*J. of Engr. for Industry*, 1967: 315-322) observa que la relación no es exacta (determinista) y que los parámetros α y β deben ser estimados a partir de los datos. Así, un modelo apropiado es el modelo multiplicativo de potencia $Y = \alpha \cdot x^{\beta} \cdot \epsilon$, que el autor ajustó a los datos que aparecen a continuación y que constan de las observaciones de la duración de 12 herramientas de carburo (tabla 13.2). Además de los valores x, y, x' y y', se dan los valores transformados pronosticados ($\hat{y}'$) y los valores pronosticados en la escala original ($\hat{y}$, después de transformar de nuevo).

Los resúmenes estadísticos para ajustar una línea recta a los datos transformados son $\sum x_i' = 74.41200$, $\sum y_i' = 26.22601$, $\sum x_i'^2 = 461.75874$, $\sum y_i'^2 = 67.74609$, y $\sum x_i' y_i' = 160.84601$ de modo que

$$\hat{\beta}_1 = \frac{160.84601 - (74.41200)(26.22601)/12}{461.75874 - (74.41200)^2/12} = -5.3996$$

$$\hat{\beta}_0 = \frac{26.22601 - (-5.3996)(74.41200)}{12} = 35.6684$$

Los valores estimados para α y β, los parámetros del modelo de función de potencia, son $\hat{\beta} = \hat{\beta}_1 = -5.3996$ y $\hat{\alpha} = e^{\hat{\beta}_0} = 3.094491530 \cdot 10^{15}$. Entonces, la función de regresión

Tabla 13.2 Datos para el ejemplo 13.3

	x	y	$x' = \ln(x)$	$y' = \ln(y)$	$\hat{y}'$	$\hat{y} = e^{\hat{y}'}$
1	600	2.35	6.39693	.85442	1.12754	3.0881
2	600	2.65	6.39693	.97456	1.12754	3.0881
3	600	3.00	6.39693	1.09861	1.12754	3.0881
4	600	3.60	6.39693	1.28093	1.12754	3.0881
5	500	6.40	6.21461	1.85630	2.11203	8.2650
6	500	7.80	6.21461	2.05412	2.11203	8.2650
7	500	9.80	6.21461	2.28238	2.11203	8.2650
8	500	16.50	6.21461	2.80336	2.11203	8.2650
9	400	21.50	5.99146	3.06805	3.31694	27.5760
10	400	24.50	5.99146	3.19867	3.31694	27.5760
11	400	26.00	5.99146	3.25810	3.31694	27.5760
12	400	33.00	5.99146	3.49651	3.31694	27.5760

estimada es $\hat{\mu}_{Y\cdot x} \approx 3.094491530 \cdot 10^{15} \cdot x^{-5.3996}$. Para recapturar la ecuación de Taylor (estimada), se establece $y = 3.094491530 \times 10^{15} \cdot x^{-5.3996}$, de donde $xy^{.185} = 740$.

La figura 13.4(a) da una gráfica de los residuos estandarizados de la regresión lineal usando variables transformadas (para las que $r^2 = .922$); no hay patrón aparente en la gráfica, aun cuando un residuo estandarizado es un poco grande, y los residuos se ven como deben para una regresión lineal simple. La figura 13.4(b) presenta una gráfica de $\hat{y}$ en función de y, que indica predicciones satisfactorias sobre la escala original.

Para obtener un intervalo de confianza para la duración mediana de la herramienta cuando el tiempo de corte es 500, se transforma $x = 500$ a $x' = 6.21461$. Entonces $\hat{\beta}_0 + \hat{\beta}_1 x' = 2.1120$, y un intervalo de confianza de 95% para $\beta_0 + \beta_1(6.21461)$ es (de la sección 12.4) $2.1120 \pm (2.228)(.0824) = (1.928, 2.296)$. El intervalo de confianza de 95% para $\tilde{\mu}_{Y\cdot 500}$ se obtiene entonces al tomar los antilogaritmos: $(e^{1.928}, e^{2.296}) = (6.876, 9.930)$. Se puede verificar con facilidad que para los datos transformados $s^2 = \hat{\sigma}^2 \approx .081$. Debido a que esto es muy pequeño, (6.876, 9.930) es un intervalo aproximado para $\mu_{Y\cdot 500}$.

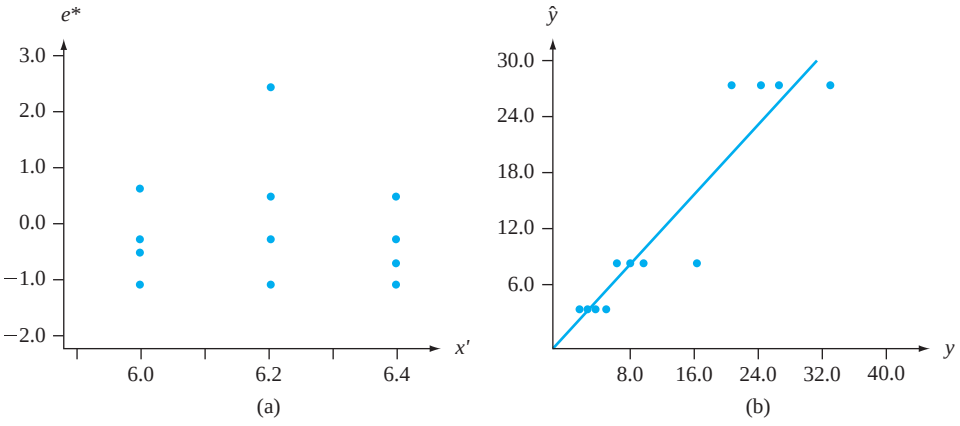


Figura 13.4 (a) Residuos estandarizados en función de x' del ejemplo 13.3; (b) $\hat{y}$ en función de y del ejemplo 13.3 ■

Ejemplo 13.4 En el artículo "Ethylene Synthesis in Lettuce Seeds: Its Physiological Significance" (*Plant Physiology*, 1972: 719-722), el contenido de etileno en semillas de lechuga (y, en nL/g peso en seco) se estudió como función de la exposición de tiempo (x, en min) a un absor-

bente de etileno. La figura 13.5 presenta un diagrama de dispersión de los datos y una gráfica de los residuos generados de una regresión lineal de y en x. Ambas gráficas muestran un fuerte patrón curvado, que sugiere que es apropiada una transformación para alcanzar linealidad. Además, una regresión lineal da predicciones negativas para $x = 90$ y $x = 100$.

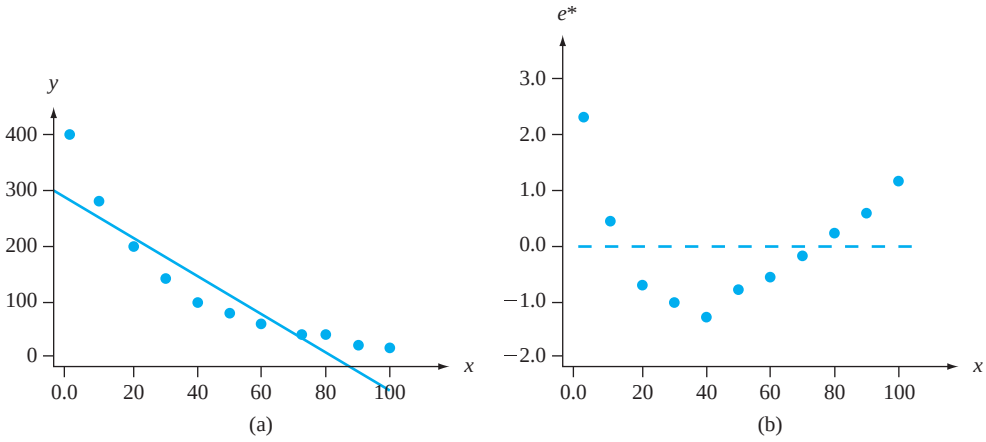


Figura 13.5 (a) Diagrama de dispersión; (b) gráfica de residuos de regresión lineal para los datos del ejemplo 13.4

El autor no dio ningún argumento para un modelo teórico, pero esta gráfica de $y' = \ln(y)$ en función de x muestra una fuerte relación lineal, que sugiere que una función exponencial dará un buen ajuste a los datos. La tabla 13.3 muestra los valores de datos y otra información de una regresión lineal de y' en x. Las estimaciones de parámetros del modelo lineal son $\hat{\beta}_1 = -.0323$ y $\hat{\beta}_0 = 5.941$, con $r^2 = .995$. La función de regresión estimada para el modelo exponencial es $\hat{\mu}_{Y \cdot x} \approx e^{\hat{\beta}_0} \cdot e^{\hat{\beta}_1 x} = 380.32e^{-.0323x}$. Los valores pronosticados $\hat{y}_i$ se pueden obtener entonces por sustitución de x_i ($i = 1, \ldots, n$) en $\hat{\mu}_{Y \cdot x}$ o bien al calcular $\hat{y}_i = e^{\hat{y}'_i}$, donde las $\hat{y}'_i$ son las predicciones del modelo transformado de línea recta. La figura 13.6 presenta una gráfica de e'^* en función de x (los residuos estandarizados de una regresión lineal) y una gráfica de $\hat{y}$ en función de y. Estas gráficas apoyan la elección de un modelo exponencial.

Tabla 13.3 Datos para el ejemplo 13.4

x	y	$y' = \ln(y)$	$\hat{y}'$	$\hat{y} = e^{\hat{y}'}$
2	408	6.01	5.876	353.32
10	274	5.61	5.617	275.12
20	196	5.28	5.294	199.12
30	137	4.92	4.971	144.18
40	90	4.50	4.647	104.31
50	78	4.36	4.324	75.50
60	51	3.93	4.001	54.64
70	40	3.69	3.677	39.55
80	30	3.40	3.354	28.62
90	22	3.09	3.031	20.72
100	15	2.71	2.708	15.00

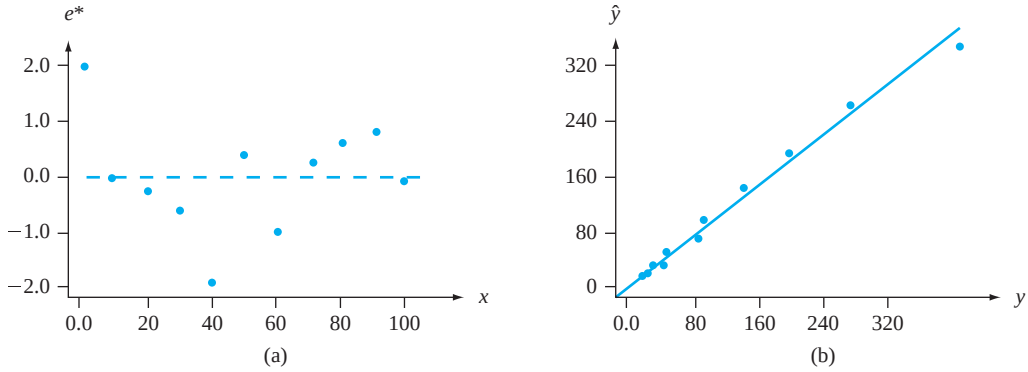


Figura 13.6 Gráfica de (a) residuos estandarizados (después de transformar) en función de x; (b) $\hat{y}$ en función de y para datos del ejemplo 13.4 ■

Al analizar datos transformados, se deben recordar los puntos siguientes:

1. Estimar β_1 y β_0 como en (13.5) y luego transformar de nuevo, para obtener estimaciones de los parámetros originales, no es equivalente a usar el principio de mínimos cuadrados directamente en el modelo original. Así, para el modelo exponencial, se podrían estimar α y β al minimizar $\sum(y_i - \alpha e^{\beta x_i})^2$. Sería necesario un cálculo iterativo. Las estimaciones resultantes no serían iguales a $\hat{\alpha} \neq e^{\hat{\beta}_0}$ y $\hat{\beta} \neq \hat{\beta}_1$.
2. Si el modelo seleccionado no es intrínsecamente lineal, el método resumido en (13.5) no se puede usar. En lugar de esto, tendrían que aplicarse mínimos cuadrados (o algún otro procedimiento de ajuste) al modelo no transformado. Por tanto, para el modelo exponencial aditivo $Y = \alpha e^{\beta x} + \epsilon$, los mínimos cuadrados comprenderían minimizar $\sum(y_i - \alpha e^{\beta x_i})^2$. Tomando derivadas parciales con respecto a α y β resulta en dos ecuaciones normales no lineales en α y β; estas ecuaciones deben resolverse entonces usando un procedimiento iterativo.
3. Cuando el modelo lineal transformado satisface todas las suposiciones mencionadas en el capítulo 12, el método de mínimos cuadrados da las mejores estimaciones de los parámetros transformados. No obstante, las estimaciones de los parámetros originales pueden no ser los mejores en ningún sentido, aun cuando serán razonables. Por ejemplo, en el modelo exponencial, el estimador $\hat{\alpha} = e^{\hat{\beta}_0}$ no será insesgado, aun cuando será el estimador de α de máxima probabilidad si el error variable ϵ' está normalmente distribuido. Usando mínimos cuadrados de manera directa (sin transformar) podría dar mejores estimaciones.
4. Si se ha hecho una transformación en y y se desea usar las fórmulas estándar para probar hipótesis o construir intervalos de confianza (IC), ϵ' debe estar distribuida al menos aproximadamente en forma normal. Para verificar esto, deben comprobarse los residuos de la regresión transformada.
5. Cuando y es transformada, el valor r^2 de la regresión resultante se refiere a una variación en las y_i' explicada por el modelo de regresión transformado. Aun cuando un alto valor de r^2 aquí indica un buen ajuste del modelo no lineal original estimado para las y_i observadas, r^2 no se refiere a las observaciones originales. Quizá la mejor forma de evaluar la calidad del ajuste es calcular los valores pronosticados $\hat{y}_i'$, usando el modelo transformado, transformándolos de nuevo a la escala y original para obtener $\hat{y}_i$, y luego trazar $\hat{y}$ en función de y. Un buen ajuste se hace entonces evidente por puntos cercanos a la recta de 45°. Se podría calcular $\text{SSE} = \sum(y_i - \hat{y}_i)^2$ como una medida numérica de la bondad de ajuste. Cuando el modelo era lineal, se comparó esto con $\text{SST} = \sum(y_i - \bar{y})^2$, la variación total alrededor de la recta horizontal a una altura $\bar{y}$, lo cual llevó a r^2. En el caso no lineal, sin embargo, no es necesariamente informativo medir la variación total en esta forma, de modo que un valor r^2 no es tan útil como en el caso lineal.

Métodos de regresión más generales

Hasta este punto se ha supuesto que $Y = f(x) + \epsilon$ (un modelo aditivo) o que $Y = f(x) \times \epsilon$ (un modelo multiplicativo). En el caso del modelo aditivo, $\mu_{Y \cdot x} = f(x)$, de modo que estimar la función de regresión $f(x)$ equivale a estimar la curva de valores medios de y. En ocasiones, un diagrama de dispersión de los datos sugiere que no hay una expresión matemática sencilla para $f(x)$. Los expertos en estadística han creado recientemente algunos métodos más flexibles que permiten modelar una amplia variedad de patrones usando el mismo procedimiento de ajuste. Uno de estos métodos es el **LOWESS** (o LOESS), que es una abreviatura de *locally weighted scatter plot smoother* (suavizador de gráfica de dispersión localmente ponderada). Denote con (x^*, y^*) uno de los n pares particulares (x, y) de la muestra. El valor $\hat{y}$ correspondiente a (x^*, y^*) se obtiene al ajustar una recta usando sólo un porcentaje especificado de los datos (por ejemplo 25%) cuyos valores x son más cercanos a x^*. Además, en lugar de usar mínimos cuadrados "ordinarios", que da un valor igual a todos los puntos, los que tienen valores x más cercanos a x^* tienen más valor que los valores x que están más alejados. La altura de la recta resultante arriba de x^* es el valor ajustado $\hat{y}^*$. Este proceso se repite para cada uno de los puntos n, de modo que n líneas diferentes se ajustan (es seguro que el lector no desearía hacer esto manualmente). Por último, los puntos ajustados se enlazan para obtener una curva LOWESS.

Ejemplo 13.5 Por lo general, no es factible pesar grandes animales muertos y encontrados en zonas silvestres, de modo que es mejor tener un método para estimar el peso a partir de diversas características de un animal que se puedan determinar con facilidad. Minitab tiene un conjunto de datos en memoria que consisten en diversas características para una muestra de $n = 143$ osos salvajes. La figura 13.7(a) muestra un diagrama de dispersión de $y =$ peso en función de $x =$ distancia alrededor del pecho (circunferencia del pecho). A primera vista, parece como si una sola recta obtenida de mínimos cuadrados ordinarios resumiera de manera eficaz el patrón. La figura 13.7(b) muestra que la curva LOWESS producida por Minitab, usando un espacio de 50% [el ajuste en (x^*, y^*), está determinado por el 50% más cercano de la muestra]. La curva parece estar formada por dos segmentos de recta unidos arriba de aproximadamente $x = 38$. La línea más inclinada a la derecha de 38, indica que el peso, tiende a aumentar con más rapidez como lo hace la circunferencia para circunferencias mayores a 38 pulgadas.

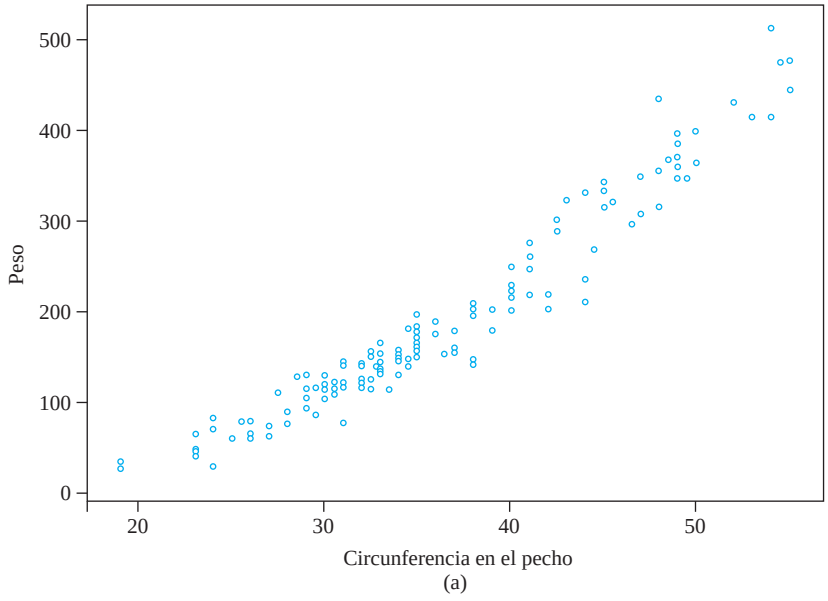


Figura 13.7 (a) Diagrama de dispersión de Minitab para datos del peso de osos

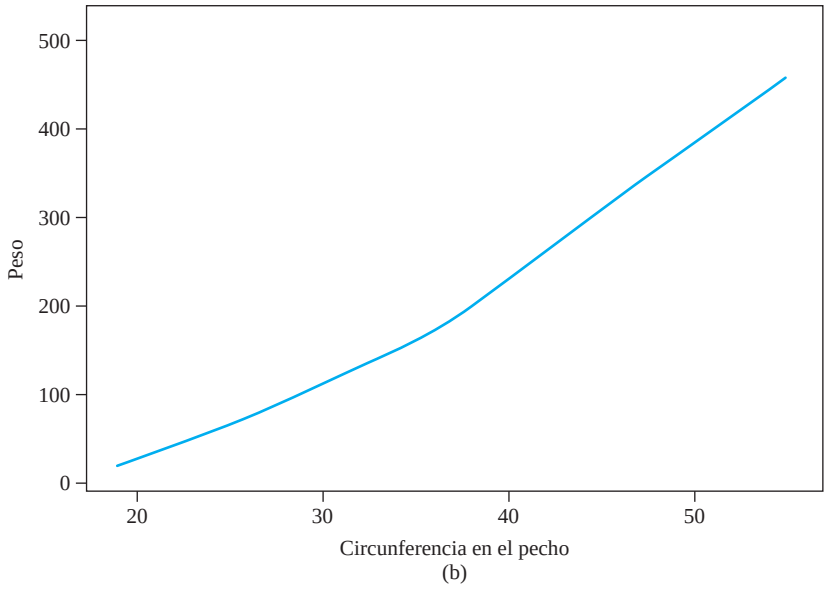


Figura 13.7 (b) Curva LOWESS de Minitab para datos del peso de osos

Es complicado hacer otras inferencias (por ejemplo, obtener un intervalo de confianza para un valor y medio) con base en este tipo general de modelo de regresión. La técnica de instrucciones preliminares mencionada antes se puede usar para este fin.

Regresión logística

El modelo sencillo de regresión lineal es apropiado para relacionar una variable cuantitativa de respuesta a un predictor cuantitativo x. Considere ahora una variable de respuesta dicotómica con valores posibles 1 y 0 correspondientes a éxito y fracaso. Sea $p = P(S) = P(y = 1)$. Con frecuencia, el valor de p dependerá del valor de alguna variable cuantitativa x. Por ejemplo, la probabilidad de que un auto necesite servicio de garantía de cierta clase podría depender de la distancia total recorrida por el vehículo, o la probabilidad de evitar una infección de cierto tipo podría depender de la dosis en una vacuna. En lugar de usar sólo el símbolo p para la probabilidad de éxito, ahora se usa $p(x)$ para resaltar la dependencia de esta probabilidad en el valor de x. La ecuación de regresión lineal simple $Y = \beta_0 + \beta_1 x + \epsilon$ ya no es apropiada, porque tomar el valor medio de cada lado de la ecuación da

$$\mu_{Y_x} = 1 \cdot p(x) + 0 \cdot (1 - p(x)) = p(x) = \beta_0 + \beta_1 x$$

Mientras que $p(x)$ es una probabilidad y por tanto debe ser entre 0 y 1, $\beta_0 + \beta_1 x$ no necesita estar en este rango.

En lugar de hacer que el valor medio de Y sea una función lineal de x, ahora se considera un modelo en el que alguna función del valor medio de Y es una función lineal de x. En otras palabras, se hace que $p(x)$ sea una función de $\beta_0 + \beta_1 x$ en lugar de $\beta_0 + \beta_1 x$ misma. Una función que se ha encontrado muy útil en numerosas aplicaciones es la **función logit**

$$p(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

La figura 13.8 muestra una gráfica de $p(x)$ para valores particulares de β_0 y β_1 con $\beta_1 > 0$. Cuando x aumenta, la probabilidad de éxito se incrementa. Para β_1 negativa, la probabilidad de éxito sería una función decreciente de x.

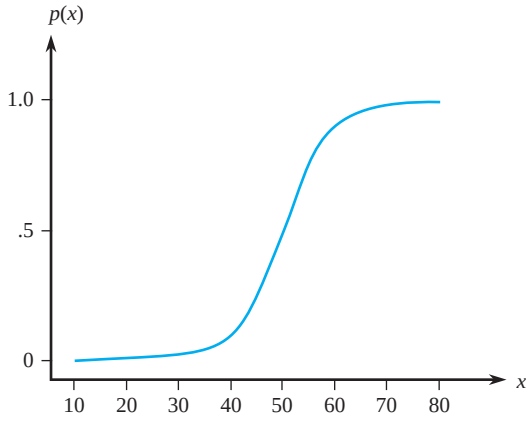


Figura 13.8 Gráfica de una función logit

Regresión logística significa suponer que $p(x)$ está relacionada a x por la función logit. Álgebra sencilla muestra que

$$\frac{p(x)}{1 - p(x)} = e^{\beta_0 + \beta_1 x}$$

La expresión del lado izquierdo recibe el nombre de *posibilidad*. Si, por ejemplo, $\frac{p(60)}{1 - p(60)} = 3$, entonces cuando $x = 60$ un éxito tiene tres veces más probabilidad que un fracaso. Ahora se ve que el logaritmo de la *posibilidad* es una función lineal del predictor. En particular, el parámetro de pendiente β_1 es el cambio en los logaritmos de las posibilidades asociadas con un aumento de 1 unidad en x. Esto implica que la posibilidad en sí cambia por el factor multiplicativo e^{β_1} cuando x aumenta en 1 unidad.

El ajuste de la regresión logística a los datos muestrales requiere que se estimen los parámetros β_0 y β_1. Por lo general esto se hace usando la técnica de máxima probabilidad descrita en el capítulo 6. Los detalles son muy complicados, pero por fortuna los paquetes de computación de estadística más populares harán esto previa solicitud y dan indicaciones cuantitativas y gráficas de qué tan bien ajusta el modelo.

Ejemplo 13.6 He aquí los datos, en forma de una pantalla comparativa de tallo y hoja, de la temperatura en marcha y la incidencia de fracaso de los empaques o juntas (O-rings) en el transbordador espacial 23 lanzamientos antes del desastre del Challenger de 1986 (S = sí, falla; N = no, no falla). Las observaciones en el lado izquierdo de la pantalla tienden a ser más pequeñas que las del lado derecho.

Sí		No	
873	5		
3	6	677789	Tallo: diez dígitos
500	7	002356689	Hoja: un dígito
	8	1	

La figura 13.9 muestra una salida Minitab para un análisis y una gráfica de la función de regresión logística estimada del software R. Se ha seleccionado denotar con p la probabilidad de falla. La gráfica de $\hat{p}$ disminuye a medida que aumenta la temperatura debido a que las fallas tienden a ocurrir a temperaturas más bajas que los éxitos. La estimación de β_1 y su desviación estándar estimada es $\hat{\beta}_1 = -.232$ y $s_{\hat{\beta}_1} = .1082$, respectivamente. Suponemos que el tamaño de muestra n es suficientemente grande aquí, así que $Z = \hat{\beta}_1 / s_{\hat{\beta}}$ tiene aproximadamente una distribución normal. Si $\beta_1 = 0$ (es decir, la temperatura no

afecta a la probabilidad de falla del empaque o junta), el estadístico de prueba $Z = \hat{\beta}_1/s_{\hat{\beta}}$ tiene aproximadamente una distribución normal estándar. El valor reportado de esta relación es de $z = -2.14$, con un valor P correspondiente de dos colas de valor P .032 (algunos paquetes reportan un valor de chi-cuadrada que está a sólo z^2, con el mismo valor P). Al nivel de significancia .05, rechazamos la hipótesis nula de no efecto de la temperatura.

Regresión logística binaria: falla contra temperatura

```
Tabla de regresión logística
                                                    Odds     95% CI
Predictor            Coef   SE Coef       Z      P  Ratio  Lower  Upper
Constante         15.0429   7.37862    2.04  0.041
temperatura     −0.232163  0.108236   −2.14  0.032   0.79   0.64   0.98

Pruebas de bondad del ajuste
Método            Chi-Cuadrada     GL         P
Pearson                11.1303     14     0.676
Desviación             11.9974     14     0.607
Hosmer-Lemeshow         9.7119      8     0.286
```

Resumen de clasificación $\hat{Y}$

Y	0	1
0	1.0000000	0.0000000
1	0.4285714	0.5714286

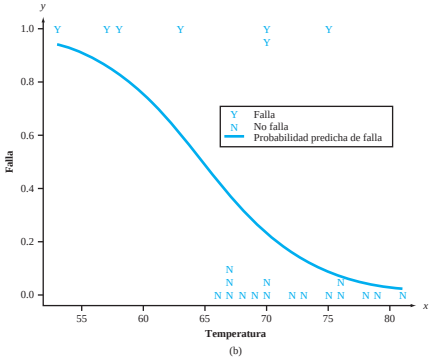


Figura 13.9 (a) Salida de regresión logística de Minitab para el ejemplo 13.6; (b) gráfica de la función logística estimada y clasificación de las probabilidades de R

Las estimaciones de las posibilidades de falla para cualquier valor particular de temperatura x es

$$\frac{p(x)}{1 - p(x)} = e^{15.0429 - .232163x}$$

Esto implica que los *cocientes de posibilidades*, las posibilidades de falla a una temperatura de $x + 1$ divididos entre las posibilidades de falla a una temperatura de x, son

$$\frac{p(x + 1)/[1 - p(x + 1)]}{p(x)/[1 - p(x)]} = e^{-.232163} = .7928$$

La interpretación es que por cada grado adicional de temperatura, se estima que las posibilidades de falla se reducirán en un factor de .79 (21%). Un IC del 95% para el cociente de posibilidad real también aparece en la salida. Además, Minitab ofrece tres diferentes formas de evaluar el modelo de falta de ajuste: las pruebas de Pearson, de desviación y de Hosmer-Lemeshow. Grandes valores P son consistentes con un buen modelo. Estas pruebas son útiles en la regresión logística múltiple, donde hay más de un predictor en el modelo de relación por lo que no es sólo una gráfica como la de la figura 13.9(b). Varias gráficas de diagnóstico están también disponibles.

La salida de R proporciona información basada en la clasificación de una observación como una falla si el $p(x)$ estimado es por lo menos .5 y como no–falla en otro caso. Puesto que $p(x) = 0.5$ cuando $x = 64.80$, tres de los siete errores (Ys en el gráfico) se clasifican erróneamente como no–fallas (una proporción de clasificación errónea de .429), mientras que ninguna de las observaciones no–falla serían mal clasificadas. Una mejor manera de evaluar la probabilidad de clasificación errónea es utilizar la *validación cruzada*: elimine la primera observación de la muestra, estime la relación, clasifique la primera observación sobre la base de esta relación de estimación y repita este proceso con cada una de las otras observaciones de la muestra (así que una observación de la muestra no afecta a su propia clasificación).

La temperatura en el lanzamiento para la misión *Challenger* era de sólo 31°F. Como este valor es mucho menor que cualquier temperatura de la muestra, es riesgoso extrapolar la relación estimada. Con todo, parece que para una temperatura así de baja, la falla de empaques es casi segura. ■

EJERCICIOS Sección 13.2 (15–25)

15. A nadie que le gusten las tortillas le gustan los pedacitos de tortilla pastosos, de modo que es importante hallar características del proceso de producción que produzcan pedacitos de tortilla con una textura atractiva. Los siguientes datos sobre x = tiempo de freír (segundos) y y = contenido de humedad (%) aparecieron en el artículo "Thermal and Physical Properties of Tortilla Chips as a Function of Frying Time" (*J. of Food Processing and Preservation*, 1995: 175–189).

x	5	10	15	20	25	30	45	60
y	16.3	9.7	8.1	4.2	3.4	2.9	1.9	1.3

a. Construya un diagrama de dispersión de y en función de x y comente.
b. Construya un diagrama de dispersión de los pares $(\ln(x), \ln(y))$ y comente.
c. ¿Qué relación probabilística entre x y y sugiere la figura lineal de la gráfica del inciso (b)?
d. Pronostique el valor del contenido de humedad cuando el tiempo de freír es 20, en una forma que lleve información acerca de la confiabilidad y precisión.
e. Analice los residuos del ajuste del modelo de regresión lineal simple a los datos transformados y comente.

16. Las cuerdas de fibra de poliéster se usan cada vez más como componentes de líneas de amarre para estructuras de mar adentro en aguas profundas. Los autores del artículo "Quantifying the Residual Creep Life of Polyester Mooring Ropes" (*Intl. J. of Offshore and Polar Exploration*, 2005: 223–228) utilizaron los datos siguientes como base para estudiar la forma en que el tiempo para falla (h) dependía de la carga (% de carga de ruptura):

x	77.7	77.8	77.9	77.8	85.5	85.5
y	5.067	552.056	127.809	7.611	.124	.077
x	89.2	89.3	73.1	85.5	89.2	85.5
y	.008	.013	49.439	.503	.362	9.930
x	89.2	85.5	89.2	82.3	82.0	82.3
y	.677	5.322	.289	53.079	7.625	155.299

Se ajustó una regresión lineal de log(tiempo) en función de la carga. Los investigadores estuvieron particularmente interesados en estimar la pendiente de la recta de regresión verdadera al relacionar estas variables. Investigue la calidad del ajuste, estime la pendiente, y pronostique el tiempo para falla cuando la carga es 80 en una forma que lleve información acerca de la confiabilidad y precisión.

17. Los datos siguientes sobre rapidez de combustión de la masa x y longitud de flama y es representativa de los que aparecieron en el artículo "Some Burning Characteristics of Filter Paper" (*Combustion Science and Technology*, 1971: 103–120):

x	1.7	2.2	2.3	2.6	2.7	3.0	3.2
y	1.3	1.8	1.6	2.0	2.1	2.2	3.0
x	3.3	4.1	4.3	4.6	5.7	6.1	
y	2.6	4.1	3.7	5.0	5.8	5.3	

a. Estime los parámetros de un modelo de función de potencia.
b. Construya gráficas de diagnóstico para verificar si una función de potencia es una opción apropiada de modelo.
c. Pruebe H_0: $\beta = \frac{4}{3}$ contra H_a: $\beta < \frac{4}{3}$, usando una prueba de nivel .05.
d. Pruebe la hipótesis nula que expresa que la longitud media de la flama, cuando la rapidez de combustión es 5.0, es el doble que la longitud media de flama cuando la rapidez de combustión es 2.5, contra la alternativa de que éste no es el caso.

18. Las fallas de turbinas de gas de aviones debidas a un alto ciclo de fatiga es un problema muy extendido. El artículo "Effect of Crystal Orientation on Fatigue Failure of Single Crystal Nickel Base Turbine Blade Superalloys (*J. of Engineering for Gas Turbines and Power*, 2002: 161–176) dio los datos siguientes y ajuste de un modelo de regresión no lineal para pronosticar la amplitud de deformación de ciclos hasta que ocurra una falla. Ajuste un modelo apropiado, investigue la calidad del ajuste, y pronostique la amplitud cuando los ciclos hasta que ocurra una falla sean = 5000.

Obs	Ciclos p/falla	Amplit. de deform.	Obs	Ciclos p/falla	Amplit. de deform.
1	1326	.01495	11	7356	.00576
2	1593	.01470	12	7904	.00580
3	4414	.01100	13	79	.01212
4	5673	.01190	14	4175	.00782
5	29516	.00873	15	34676	.00596
6	26	.01819	16	114789	.00600
7	843	.00810	17	2672	.00880
8	1016	.00801	18	7532	.00883
9	3410	.00600	19	30220	.00676
10	7101	.00575			

19. Se realizaron pruebas de resistencia térmica para estudiar la relación entre temperatura y duración de alambre esmaltado de poliéster ("Thermal Endurance of Polyester Enameled Wires Using Twisted Wire Specimens", *IEEE Trans. Insulation*, 1965: 38–44), que dieron por resultado los datos siguientes.

Temp.	200	200	200	200	200	200
Duración	5933	5404	4947	4963	3358	3878
Temp.	220	220	220	220	220	220
Duración	1561	1494	747	768	609	777
Temp.	240	240	240	240	240	240
Duración	258	299	209	144	180	184

a. Un diagrama de dispersión de los datos ¿sugiere una relación probabilística lineal entre duración y temperatura?
b. ¿Qué modelo está implicado por una relación lineal entre ln(duración) esperada y 1/temperatura? ¿Aparece un diagrama de dispersión de los datos transformados consistente con esta relación?
c. Estime los parámetros del modelo sugerido en el inciso (b). ¿Qué duración se pronosticaría para una temperatura de 220?
d. Debido a que hay múltiples observaciones en cada valor x, el método del ejercicio 14 se puede usar para probar la hipótesis nula que expresa que el modelo sugerido en el inciso (b) es correcto. Realice la prueba al nivel .01.

20. El ejercicio 14 presentó datos sobre el peso corporal x y la rapidez de eliminación metabólica/peso corporal y. Considere las siguientes funciones intrínsecamente lineales para especificar la relación entre las dos variables: (a) $\ln(y)$ en función de x, (b) $\ln(y)$ en función de $\ln(x)$, (c) y en función de $\ln(x)$, (d) y en función de $1/x$, y (e) $\ln(y)$ en función de $1/x$. Use cualesquiera gráficas de diagnóstico apropiadas y análisis para decidir cuáles de estas funciones seleccionaría para especificar un modelo probabilístico. Explique su razonamiento.

21. Una gráfica del artículo "Thermal Conductivity of Polyethylene: The Effects of Crystal Size, Density, and Orientation on the Thermal Conductivity" (*Polymer Engr. and Science*, 1972: 204–208) sugiere que el valor esperado de conductividad térmica y es una función lineal de $10^4 \cdot 1/x$, donde x es el grosor laminar.

x	240	410	460	490	520	590	745	8300
y	12.0	14.7	14.7	15.2	15.2	15.6	16.0	18.1

a. Estime los parámetros de la función de regresión y la función de regresión en sí.
b. Pronostique el valor de conductividad térmica cuando el grosor laminar sea de 500 Å.

22. En cada uno de los casos siguientes, decida si la función dada es intrínsecamente lineal. Si es así, identifique x' y y', y luego explique la forma en que un término de error aleatorio ϵ se puede introducir para dar un modelo probabilístico intrínsecamente lineal.
a. $y = 1/(\alpha + \beta x)$
b. $y = 1/(1 + e^{\alpha+\beta x})$
c. $y = e^{e^{\alpha+\beta x}}$ (una curva Gompertz)
d. $y = \alpha + \beta e^{\lambda x}$

23. Suponga que x y y están relacionadas de acuerdo con un modelo exponencial probabilístico $Y = \alpha e^{\beta x} \cdot \epsilon$, con $V(\epsilon)$ una constante independiente de x (como fue el caso en el modelo lineal sencillo $Y = \beta_0 + \beta_1 x + \epsilon$). ¿Es $V(Y)$ una constante independiente de x [como fue el caso para $Y = \beta_0 + \beta_1 x + \epsilon$, donde $V(Y) = \sigma^2$]? Explique su razonamiento. Trace una figura de un diagrama de dispersión prototipo que resulte de este modelo. Conteste las mismas preguntas para el modelo de potencia $Y = \alpha x^{\beta} \cdot \epsilon$.

24. La cifosis es una grave flexión hacia adelante de la espina dorsal que se presenta después de una cirugía espinal correctiva. Un estudio realizado para determinar factores de riesgo por la cifosis informó de las edades siguientes (meses) de 40 personas en el momento de la operación; las primeras 18 personas tenían cifosis, no así las 22 restantes.

Con cifosis	12	15	42	52	59	73
	82	91	96	105	114	120
	121	128	130	139	139	157
Sin cifosis	1	1	2	8	11	18
	22	31	37	61	72	81
	97	112	118	127	131	140
	151	159	177	206		

Utilice la regresión logística generada por Minitab que aparece en la página 543, para determinar si la edad parece tener un impacto significativo en la presencia de cifosis.

25. El artículo "Acceptable Noise Levels for Construction Site Offices" (*Building Serv. Engr. Res. Tech.*, 2009: 87–94) analizó las respuestas de una muestra de 77 individuos, a cada uno de los cuales se le pidió decir si un nivel de ruido en particular (dBA) al que había estado expuesto es aceptable o inaceptable. He aquí los datos proporcionados por los autores del artículo:

Tabla de regresión logística para el ejercicio 24

Predictor	Coef	DE Coef	Z	P	Cociente de posibilidad	95% IC Inferior	95% IC Superior
Constante	−0.5727	0.6024	−0.95	0.342			
edad	0.004296	0.005849	0.73	0.463	1.00	0.99	1.02

Tabla de regresión logística para el ejercicio 25

Predictor	Coef	SE Coef	Z	P	Cociente de posibilidad	95% IC Inferior	95% IC Superior
Constante	23.2124	5.05095	4.60	0.000			
nivel de ruido	−0.359441	0.0785031	−4.58	0.000	0.70	0.60	0.81

Aceptable:

55.3 55.3 55.3 55.9 55.9 55.9 55.9 56.1 56.1 56.1 56.1
56.1 56.1 56.8 56.8 57.0 57.0 57.0 57.8 57.8 57.8 57.9
57.9 57.9 58.8 58.8 58.8 59.8 59.8 59.8 62.2 62.2 65.3
65.3 65.3 65.3 68.7 69.0 73.0 73.0

Inaceptable:

63.8 63.8 63.8 63.9 63.9 63.9 64.7 64.7 64.7 65.1 65.1
65.1 67.4 67.4 67.4 67.4 68.7 68.7 68.7 70.4 70.4 71.2
71.2 73.1 73.1 74.6 74.6 74.6 74.6 79.3 79.3 79.3 79.3
79.3 83.0 83.0 83.0

Interprete la regresión logística generada por Minitab y trace una gráfica de la probabilidad estimada del nivel aceptable de ruido como función del nivel.

13.3 Regresión polinomial

Los modelos no lineales, pero intrínsecamente lineales de la sección 13.2, comprendían funciones de la variable independiente x que eran estrictamente crecientes o estrictamente decrecientes. En numerosas situaciones, ya sea de un razonamiento teórico o de otro tipo, un diagrama de dispersión de los datos sugiere que la verdadera función de regresión $\mu_{Y \cdot x}$ tiene uno o más picos o valles, es decir, al menos un mínimo o máximo relativos. En tales casos, una función con polinomios $y = \beta_0 + \beta_1 x + \cdots + \beta_k x^k$ puede dar una aproximación satisfactoria a la verdadera función de regresión.

DEFINICIÓN

La **ecuación del modelo de regresión con polinomios de *k*-ésimo grado** es

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_k x^k + \epsilon \qquad (13.6)$$

donde ϵ es una variable aleatoria distribuida normalmente con

$$\mu_\epsilon = 0 \qquad \sigma_\epsilon^2 = \sigma^2 \qquad (13.7)$$

De (13.6) y (13.7), se deduce de inmediato que

$$\mu_{Y \cdot x} = \beta_0 + \beta_1 x + \cdots + \beta_k x^k \qquad \sigma_{Y \cdot x}^2 = \sigma^2 \qquad (13.8)$$

En otras palabras, el valor esperado de Y es una función con polinomios de k-ésimo grado de x, mientras que la varianza de Y, que controla la dispersión de valores observados alrededor de la función de regresión, es la misma para cada valor de x. Se supone que los pares observados $(x_1, y_1), \ldots, (x_n, y_n)$ se generaron de manera independiente del modelo (13.6). La figura 13.10 ilustra un modelo cuadrático y uno cúbico; en la práctica, es muy raro ir más allá de $k = 3$.

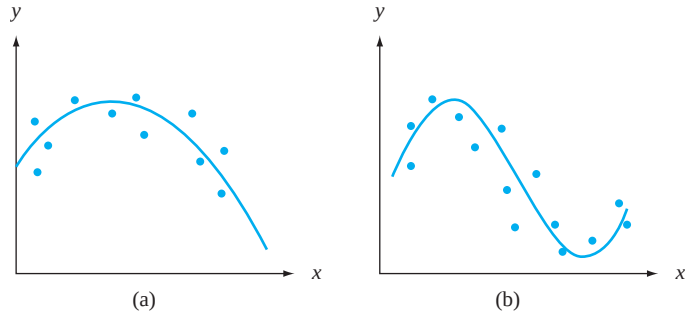


Figura 13.10 (a) Modelo de regresión cuadrático; (b) modelo de regresión cúbico

Estimación de parámetros

Para estimar las β considere una función de regresión de prueba $y = b_0 + b_1x + \cdots + b_kx^k$. Entonces la bondad de ajuste de esta función a los datos observados se puede evaluar al calcular la suma de desviaciones al cuadrado

$$f(b_0, b_1, \ldots, b_k) = \sum_{i=1}^{n}[y_i - (b_0 + b_1x_i + b_2x_i^2 + \cdots + b_kx_i^k)]^2 \qquad (13.9)$$

Según el principio de mínimos cuadrados, las estimaciones $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ son los valores de $b_0, b_1, \ldots, b_k$ que minimizan la expresión (13.9). Debe observarse que cuando $x_1, x_2, \ldots, x_n$ son todas diferentes, hay un polinomio de grado $n - 1$ que se ajusta a los datos perfectamente, de modo que el valor minimizador de (13.9) es 0 cuando $k = n - 1$. No obstante, en casi todas las aplicaciones, el modelo con polinomios (13.6) con k grande es bastante irreal.

Para hallar los valores minimizadores en (13.9), se toman las $k + 1$ derivadas parciales $\partial f/\partial b_0, \partial f/\partial b_1, \ldots, \partial f/\partial b_k$ y se igualan a 0, lo cual produce el sistema de ecuaciones normales para las estimaciones. Debido a que la función de prueba $b_0 + b_1x + \cdots + b_kx^k$ es lineal en $b_0, \ldots, b_k$ (aunque no en x), las $k + 1$ ecuaciones normales son lineales en las incógnitas:

$$\begin{aligned} b_0n + b_1\Sigma x_i + b_2\Sigma x_i^2 + \cdots + b_k\Sigma x_i^k &= \Sigma y_i \\ b_0\Sigma x_i + b_1\Sigma x_i^2 + b_2\Sigma x_i^3 + \cdots + b_k\Sigma x_i^{k+1} &= \Sigma x_iy_i \\ \vdots \qquad\quad \vdots \qquad\quad & \vdots \\ b_0\Sigma x_i^k + b_1\Sigma x_i^{k+1} + \cdots + b_k\Sigma x_i^{2k} &= \Sigma x_i^ky_i \end{aligned} \qquad (13.10)$$

Todos los paquetes de computadora estándares de estadística resolverán de manera automática las ecuaciones de (13.10) y darán las estimaciones junto con otra gran cantidad de información.*

Ejemplo 13.7 El artículo "Residual Stresses and Adhesion of Thermal Spray Coatings" (*Surface Engineering*, 2005: 35–40) consideró la relación entre el grosor (μm) de capas de NiCrAl depositadas en sustrato de acero inoxidable y la resistencia a la adherencia (MPa). Los datos que aparecen a continuación se interpretaron de una gráfica del artículo citado antes.

Grosor	220	220	220	220	370	370	370	370	440	440
Resistencia	24.0	22.0	19.1	15.5	26.3	24.6	23.1	21.2	25.2	24.0
Grosor	440	440	680	680	680	680	860	860	860	860
Resistencia	21.7	19.2	17.0	14.9	13.0	11.8	12.2	11.2	6.6	2.8

*En la sección 13.4 se estudia que la regresión con polinomios es un caso especial de regresión múltiple, de modo que en general se usa un comando apropiado para este último trabajo.

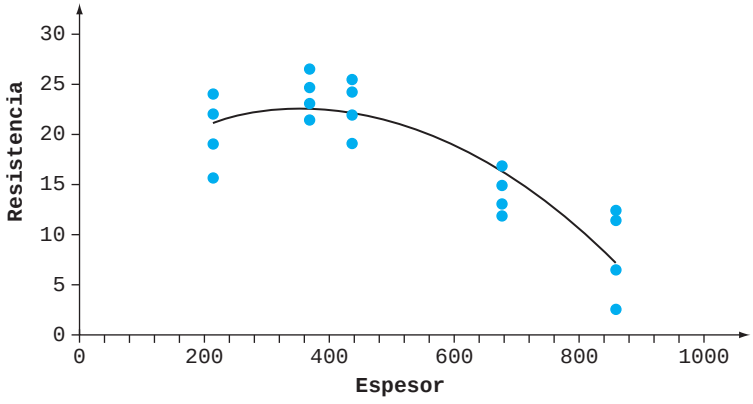
El diagrama de dispersión de la figura 13.11(a) apoya la selección del modelo de regresión cuadrático. La figura 13.11(b) contiene una salida de Minitab de un ajuste de este modelo. Los coeficientes estimados de regresión son

$$\hat{\beta}_0 = 14.521 \qquad \hat{\beta}_1 = .04323 \qquad \hat{\beta}_2 = -.00006001$$

de los cuales la función estimada de regresión es

$$y = 14.521 + .04323x - .00006001x^2$$

La sustitución de los valores de x sucesivos 220, 220, . . . , 860, y 860 en esta función da los valores pronosticados $\hat{y}_1 = 21.128, \ldots, \hat{y}_{20} = 7.321$ y los residuos $y_1 - \hat{y}_1 = 2.872, \ldots, y_{20} - \hat{y}_{20} = -4.521$ resultan de la sustracción. La figura 13.12 muestra una gráfica de los residuos estandarizados en función de $\hat{y}$ y también una gráfica de probabilidad normal de los residuos estandarizados, los cuales validan el modelo cuadrático.




```
La ecuación de regresión es
Resistencia = 14.5 + 0.0432 espesor − 0.000060 espesor al cuadrado

Predictor                      Coef        SE Coef          T          P
Constante                    14.521          4.754       3.05      0.007
espesor                     0.04323        0.01981       2.18      0.043
espesor al cuadrado      −0.00006001     0.00001786      −3.36      0.004

S = 3.26937      R-Sq = 78.0%      R-Sq(adj) = 75.4%

Análisis de varianza

Fuente              GL          SS          MS          F          P
Regresión            2      643.29      321.65      30.09      0.000
Error residual      17      181.71       10.69
Total               19      825.00

Valores pronosticados para nuevas observaciones

Nuevos
Obs      Ajuste    SE Ajustada        95% IC              95% IP
  1      21.136          1.167   (18.674, 23.598)    (13.812, 28.460)
  2      10.704          1.189   ( 8.195, 13.212)    ( 3.364, 18.043)

Values of Predictors for New Observations

Nuevas
Obs              espesor      espesor al cuadrado
  1                  500                   250000
  2                  800                   640000
```

Figura 13.11 Diagrama de dispersión de datos del ejemplo 13.7 y salida Minitab del ajuste del modelo cuadrático

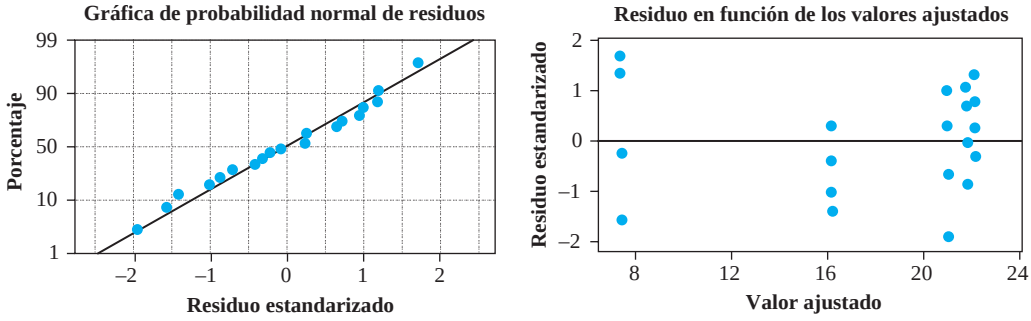


Figura 13.12 Gráficas de diagnóstico para ajuste de modelo cuadrático a datos del ejemplo 13.7 ■

$\hat{\sigma}^2$ y R^2

Para hacer inferencias adicionales, debe estimarse la varianza de error σ^2. Con $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \cdots + \hat{\beta}_k x_i^k$, el residuo i-ésimo es $y_i - \hat{y}_i$ y la suma del cuadrado de residuos (suma de errores cuadráticos) es $\text{SSE} = \sum(y_i - \hat{y}_i)^2$. La estimación de σ^2 es entonces

$$\hat{\sigma}^2 = s^2 = \frac{\text{SSE}}{n - (k + 1)} = \text{MSE} \tag{13.11}$$

donde el denominador $n - (k + 1)$ se usa porque $k + 1$ grados de libertad se pierden al estimar $\beta_0, \beta_1, \ldots, \beta_k$.

Si de nuevo se hace $\text{SST} = \sum(y_i - \bar{y})^2$, entonces SSE/SST es la proporción de la variación total en las y_i observadas que no es explicada por el modelo polinomial. La cantidad $1 - \text{SSE/SST}$, la proporción de variación explicada por el modelo, recibe el nombre de **coeficiente de determinación múltiple** y se denota con R^2.

Considere ajustar un modelo cúbico a los datos del ejemplo 13.7. Debido a que el modelo cúbico incluye el cuadrático como un caso especial, el ajuste de un cúbico será al menos tan bueno como el ajuste a un cuadrático. En forma más general, con SSE_k = suma de los errores cuadráticos de un polinomio de k-ésimo grado, $\text{SSE}_{k'} \leq \text{SSE}_k$ y $R^2_{k'} \geq R^2_k$ siempre que $k' > k$. Como el objetivo del análisis de regresión es hallar un modelo que sea sencillo (con relativamente pocos parámetros) y que dé un buen ajuste a los datos, un polinomio de grado superior puede no especificar un modelo mejor que un modelo de grado inferior a pesar de su mayor valor de R^2. Para equilibrar el costo de usar más parámetros contra la ganancia en R^2, muchos expertos en estadística usan el **coeficiente ajustado de determinación múltiple**

$$R^2 \text{ ajustada} = 1 - \frac{n - 1}{n - (k + 1)} \cdot \frac{\text{SSE}}{\text{SST}} = \frac{(n - 1)R^2 - k}{n - 1 - k} \tag{13.12}$$

La R^2 ajustada adecua hacia arriba la proporción de variación no explicada [porque la razón $(n - 1)/(n - k - 1)$ excede de 1], que resulta en R^2 *ajustada* $< R^2$. Entonces, si $R_2^2 = .66$, $R_3^2 = .70$, y $n = 10$, entonces

$$\text{ajustada } R_2^2 = \frac{9(.66) - 2}{10 - 3} = .563 \quad \text{ajustada } R_3^2 = \frac{9(.70) - 3}{10 - 4} = .550$$

de modo que la pequeña ganancia en R^2, al pasar de un modelo cuadrático a uno cúbico, no es suficiente para compensar el costo de sumar un parámetro extra al modelo.

Ejemplo 13.8 (Continuación del ejemplo 13.7)

SSE y SST se encuentran por lo general en salidas de computadora en una tabla ANOVA. La figura 13.11(b) da $\text{SSE} = 181.71$ y $\text{SST} = 825.00$, para los datos de resistencia a la adherencia, de donde $R^2 = 1 - 181.71/825.00 = .780$ (alternativamente, $R^2 = \text{SSR/SST} = 643.29/825.00 = .780$). Así, 78.0% de la variación observada en resistencia a la adhe-

rencia se puede atribuir a la relación del modelo. R^2 ajustada $= .754$, es sólo un pequeño cambio hacia abajo en R^2. Las estimaciones de σ^2 y σ son

$$\hat{\sigma}^2 = s^2 = \frac{\text{SSE}}{n-(k+1)} = \frac{181.71}{20-(2+1)} = 10.69$$

$$\hat{\sigma} = s = 3.27$$

■

Además de calcular R^2 y ajustar R^2, se deben examinar las gráficas de diagnóstico usuales para determinar si son válidas las suposiciones del modelo o si puede ser apropiada una modificación (vea la figura 13.12). También existe una prueba formal de modelos de utilidad, una prueba de F con base en las sumas de cuadrados ANOVA. Dado que la regresión polinómica es un caso especial de regresión múltiple, aplazaremos el debate de esta prueba hasta la siguiente sección.

Intervalos estadísticos y procedimientos de prueba

Debido a que las y_i aparecen en las ecuaciones normales (13.10) sólo en el lado derecho y en forma lineal, las estimaciones resultantes $\hat{\beta}_0, \ldots, \hat{\beta}_k$ son por sí mismas funciones lineales de las y_i. En esta forma, los estimadores son funciones lineales de las Y_i, de modo que cada $\hat{\beta}_i$ tiene una distribución normal. También se puede demostrar que cada $\hat{\beta}_i$ es un estimador insesgado de β_i.

La desviación estándar del estimador $\hat{\beta}_i$ se denota con $\sigma_{\hat{\beta}_i}$. Esta desviación estándar tiene la forma

$$\sigma_{\hat{\beta}_i} = \sigma \cdot \left\{ \begin{array}{l} \text{expresión complicada que comprende } \textit{todas} \text{ las} \\ x_j\text{'s, } x_j^2\text{'s, } \ldots, \text{ y } x_j^k\text{'s} \end{array} \right\}$$

Por fortuna, la expresión dentro de llaves se ha programado en todos los paquetes computarizados de estadística que se usan con más frecuencia. La desviación estándar estimada de $\hat{\beta}_i$, resulta de sustituir s en lugar de σ en la expresión para $\sigma_{\hat{\beta}_i}$. Estas desviaciones estándar estimadas $s_{\hat{\beta}_0}, s_{\hat{\beta}_1}, \ldots,$ y $s_{\hat{\beta}_k}$ aparecen en la salida de todos los paquetes de estadística citados líneas antes. Se denota con $S_{\hat{\beta}_i}$ el estimador de $\sigma_{\hat{\beta}_i}$, es decir, la variable aleatoria cuyo valor observado es $s_{\hat{\beta}_i}$. Entonces se puede demostrar que la variable estandarizada

$$T = \frac{\hat{\beta}_i - \hat{\beta}_i}{S_{\hat{\beta}_i}} \tag{13.13}$$

tiene una distribución t basada en $n-(k+1)$ grados de libertad. Esto lleva a los procedimientos inferenciales siguientes.

Un intervalo de confianza (IC) $100(1-\alpha)\%$ para β_i, el coeficiente de x^i de la función de regresión con polinomios, es

$$\hat{\beta}_i \pm t_{\alpha/2,n-(k+1)} \cdot s_{\hat{\beta}_i}$$

Una prueba de H_0: $\beta_i = \beta_{i0}$ está basada en el valor del estadístico t

$$t = \frac{\hat{\beta}_i - \beta_{i0}}{s_{\hat{\beta}_i}}$$

La prueba está basada en $n-(k+1)$ grados de libertad y es de cola superior, cola inferior, o de dos colas, según si la desigualdad en H_a es $>$, $<$ o $\neq$.

Una estimación puntual de $\mu_{Y \cdot x}$, es decir, de $\beta_0 + \beta_1 x + \cdots + \beta_k x^k$, es $\hat{\mu}_{Y \cdot x} = \hat{\beta}_0 + \hat{\beta}_1 x + \cdots + \hat{\beta}_k x^k$. La desviación estándar estimada del estimador correspondiente es más bien complicada. Numerosos paquetes de computadora darán esta desviación estándar

estimada para *cualquier* valor de x cuando un usuario lo pida. Esto, junto con una variable t estandarizada apropiada, se puede usar para justificar los procedimientos siguientes.

Se denota con x^* un valor específico de x. Un intervalo de confianza (IC) $100(1-\alpha)\%$ para $\mu_{Y\cdot x^*}$ es

$$\hat{\mu}_{Y\cdot x^*} \pm t_{\alpha/2,n-(k+1)} \cdot \left\{ \begin{matrix} \text{desv. est. estimada de} \\ \hat{\mu}_{Y\cdot x^*} \end{matrix} \right\}$$

Con $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x^* + \cdots + \hat{\beta}_k (x^*)^k$, $\hat{y}$ denotando el valor calculado de $\hat{Y}$ para los datos dados y $s_{\hat{Y}}$ denotando la desviación estándar estimada de la estadística $\hat{Y}$, la fórmula para el intervalo de confianza es muy semejante a la del caso de regresión lineal simple:

$$\hat{y} \pm t_{\alpha/2,n-(k+1)} \cdot s_{\hat{Y}}$$

Un intervalo de predicción $100(1-\alpha)\%$ para un valor y futuro a observar cuando $x = x^*$ es

$$\hat{\mu}_{Y\cdot x^*} \pm t_{\alpha/2,n-(k+1)} \cdot \left\{ s^2 + \left(\begin{matrix} \text{desv. est.} \\ \text{estimada de } \hat{\mu}_{Y\cdot x^*} \end{matrix} \right)^2 \right\}^{1/2} = \hat{y} \pm t_{\alpha/2,n-(k+1)} \cdot \sqrt{s^2 + s_{\hat{Y}}^2}$$

Ejemplo 13.9 (Continuación del ejemplo 13.8)

La figura 13.11(b) muestra que $\hat{\beta}_2 = -.00006001$ y $s_{\hat{\beta}_2} = .00001786$ (de la columna de coeficientes SE al principio de la salida). La hipótesis nula H_0: $\beta_2 = 0$ dice que mientras el predictor lineal x se retenga en el modelo, el predictor cuadrático x^2 no proporciona información útil adicional. La alternativa relevante es H_a: $\beta_2 \neq 0$ y el estadístico de prueba es $T = \hat{\beta}_2 / S_{\hat{\beta}_2}$, con valor calculado -3.36. La prueba está basada en $n - (k+1) = 17$ grados de libertad. Al nivel de significancia de .05, la hipótesis nula es rechazada porque $-3.36 \leq -2.110 = -t_{.025,\,17}$. La inclusión del predictor cuadrático se justifica. La misma conclusión resulta de comparar el valor P reportado de .004 al nivel escogido de significación de .05.

La salida de la figura 13.11(b) también contiene información de estimación y predicción para $x = 500$ y para $x = 800$. En particular, para $x = 500$,

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1(500) + \hat{\beta}_2(500)^2 = \text{Ajuste} = 21.136$$
$$s_{\hat{Y}} = \text{desv. est. estimada de } \hat{Y} = \text{Ajuste SE} = 1.167$$

de la cual un intervalo de confianza de 95% para resistencia media cuando el grosor es $= 500$ es $21.136 \pm (2.110) \times (1.167) = (18.67, 23.60)$. Un intervalo de predicción de 95% para la resistencia que resulta de una sola adherencia cuando el grosor es $= 500$ es $21.136 \pm (2.110)[(3.27)^2 + (1.167)^2]^{1/2} = (13.81, 28.46)$. Como ya se dijo antes, el intervalo de predicción es considerablemente más ancho que el intervalo de confianza porque s es grande en comparación con el ajuste SE. ■

Centrado de valores x

Para el modelo cuadrático con función de regresión $\mu_{Y\cdot x} = \beta_0 + \beta_1 x + \beta_2 x^2$, los parámetros β_0, β_1 y β_2 caracterizan el comportamiento de la función cerca de $x = 0$. Por ejemplo, β_0 es la altura a la que la función de regresión cruza el eje vertical $x = 0$, mientras que β_1 es la primera derivada de la función en $x = 0$ (rapidez de cambio instantánea de $\mu_{Y\cdot x}$ en $x = 0$). Si todas las x_i están lejos de 0, no se puede tener información precisa acerca de los valores de estos parámetros. Sea $\bar{x}$ = promedio de las x_i para las que se toman observaciones, y considere el modelo

$$Y = \beta_0^* + \beta_1^*(x - \bar{x}) + \beta_2^*(x - \bar{x})^2 + \epsilon \qquad (13.14)$$

En el modelo (13.14), $\mu_{Y \cdot x} = \beta_0^* + \beta_1^*(x - \bar{x}) + \beta_2^*(x - \bar{x})^2$ y los parámetros ahora describen el comportamiento de la función de regresión cerca del centro $\bar{x}$ de los datos.

Para estimar los parámetros de (13.14), simplemente se resta $\bar{x}$ de cada x_i para obtener $x_i' = x_i - \bar{x}$, y luego se usan las x_i' en lugar de las x_i. Un beneficio importante de esto es que los coeficientes de $b_0, \ldots, b_k$ de las ecuaciones normales (13.10) serán de magnitud mucho menor de lo que sería el caso si se usaran las x_i originales. Cuando el sistema se resuelve en computadora, este centrado protege contra cualquier error de redondeo que pueda resultar.

Ejemplo 13.10 El artículo "A Method for Improving the Accuracy of Polynomial Regression Analysis" (*J. of Quality Tech.*, 1971: 149-155) informa acerca de los datos siguientes sobre x = temperatura de cura (°F) y y = resistencia máxima al corte de un compuesto de caucho (en libras por pulgada cuadrada), con $\bar{x} = 297.13$:

x	280	284	292	295	298	305	308	315
x'	−17.13	−13.13	−5.13	−2.13	.87	7.87	10.87	17.87
y	770	800	840	810	735	640	590	560

Un análisis de computadora dio los resultados que se ilustran en la tabla 13.4.

Tabla 13.4 Coeficientes estimados y desviaciones estándar para el ejemplo 13.10

Parámetro	Estimado	Desv. est. est.	Parámetro	Estimado	Desv. Est. est.
β_0	−26,219.64	11,912.78	β_0^*	759.36	23.20
β_1	189.21	80.25	β_1^*	−7.61	1.43
β_2	−.3312	.1350	β_2^*	−.3312	.1350

La función de regresión estimada usando el modelo original es $y = -26{,}219.64 + 189.21x - .3312x^2$, mientras que el modelo centrado de la función es $y = 759.36 - 7.61(x - 297.13) - .3312(x - 297.13)^2$. Estas funciones estimadas son idénticas; la única diferencia es que se han estimado parámetros diferentes para los dos modelos. Las desviaciones estándar estimadas indican con claridad que β_0^* y β_1^* se han estimado con más precisión que β_0 y β_1. Los parámetros cuadráticos son idénticos ($\beta_2 = \beta_2^*$), como se puede ver al comparar el término x^2 en (13.14) con el modelo original. Otra vez se destaca aquí que un beneficio importante del centrado es la ganancia en precisión computacional, no sólo en modelos cuadráticos sino también de orden superior. ■

El libro de Neter y otros, que aparece en la bibliografía del capítulo, es una buena fuente de más información acerca de una regresión con polinomios.

EJERCICIOS Sección 13.3 (26–35)

26. El artículo "Physical Properties of Cumin Seed" (*J. of Agric. Engr. Res.*, 1996: 93–98) considera una regresión cuadrática de y = densidad a granel en x = contenido de humedad. Los datos de un gráfico en el artículo siguiente, junto con la salida de Minitab del ajuste cuadrático.

```
La ecuación de regresión es
densidad a granel = 403 + 16.2 contenido de humedad
− 0.706 contenido al cuadrado

Predictor          Coef        DE         T         P
Constante        403.24     36.45     11.06     0.002
contenido
de humedad       16.164     5.451      2.97     0.059
contenido
al cuadrado     −0.7063    0.1852     −3.81     0.032

S = 10.15  R-cuadrado = 93.8%  R-cuadrado(adj) = 89.6%

Análisis de varianza

Fuente            GL       SS       MS       F       P
Regresión          2   4637.7   2318.9   22.51   0.016
Error residual     3    309.1    103.0
Total              5   4946.8

      contenido densidad                      DE        St
Obs  de humedad a granel  Ajuste  Ajuste  Residual  Resid
 1        7.0     479.00  481.78   9.35    −2.78    −0.70
 2       10.3     503.00  494.79   5.78     8.21     0.98
 3       13.7     487.00  492.12   6.49    −5.12    −0.66
 4       16.6     470.00  476.93   6.10    −6.93    −0.85
 5       19.8     458.00  446.39   5.69    11.61     1.38
 6       22.0     412.00  416.99   8.75    −4.99    −0.97

              DE
Ajuste    Ajuste      95.0% IC            95.0% IP
491.10     6.52   (470.36, 511.83)   (452.71, 529.48)
```

a. El diagrama de dispersión de los datos ¿parece consistente con el modelo de regresión cuadrático?

b. ¿Qué proporción de variación observada en densidad se puede atribuir a la relación del modelo?

c. Calcule un IC de 95% para el promedio de densidad real cuando el contenido de humedad es 13.7.

d. El último renglón de la salida es de una información de petición de estimación y predicción cuando el contenido de humedad es 14. Calcule un intervalo de predicción de 99% para densidad cuando el contenido de humedad sea de 14.

e. ¿El predictor cuadrático parece dar información útil? Pruebe las hipótesis apropiadas al nivel .05 de significación.

27. Los datos siguientes sobre y = concentración de glucosa (g/L), y x = tiempo de fermentación (días) para una mezcla particular de licor de malta, se leyeron de un diagrama de dispersión del artículo "Improving Fermentation Productivity with Reverse Osmosis" (*Food Tech.*, 1984: 92–96):

x	1	2	3	4	5	6	7	8
y	74	54	52	51	52	53	58	71

a. Verifique que un diagrama de dispersión de los datos sea consistente con la selección de un modelo de regresión cuadrático.

b. La ecuación de regresión cuadrática estimada es $y = 84.482 - 15.875x + 1.7679x^2$. Pronostique el valor de concentración de glucosa para un tiempo de fermentación de 6 días, y calcule el residuo correspondiente.

c. Usando SSE = 61.77, ¿qué proporción de variación observada se puede atribuir a la relación de regresión cuadrática?

d. Los $n = 8$ residuos estandarizados basados en el modelo cuadrático son 1.91, −1.95, −.25, .58, .90, .04, −.66, y .20. Construya una gráfica de los residuos estandarizados contra x y una gráfica de probabilidad normal. ¿Estas gráficas muestran algunas características problemáticas?

e. La desviación estándar estimada de $\hat{\mu}_{Y\cdot 6}$, es decir, $\hat{\beta}_0 + \hat{\beta}_1(6) + \hat{\beta}_2(36)$, es 1.69. Calcule un intervalo de confianza de 95% para $\mu_{Y\cdot 6}$.

f. Calcule un intervalo de predicción de 95% para una observación de concentración de glucosa hecha después de 6 días de tiempo de fermentación.

28. La viscosidad (y) de un aceite se midió con un cono y un viscosímetro de plato a seis velocidades de cono diferentes (x). Se supuso que un modelo de regresión cuadrático era apropiado, y la función de regresión estimada resultante de las $n = 6$ observaciones fue

$$y = -113.0937 + 3.3684x - .01780x^2$$

a. Estime $\mu_{Y\cdot 75}$, la viscosidad esperada cuando la velocidad es 75 rpm.

b. ¿Qué viscosidad se pronosticaría para una velocidad de cono de 60 rpm?

c. Si $\sum y_i^2 = 8386.43$, $\sum y_i = 210.70$, $\sum x_i y_i = 17{,}002.00$, y $\sum x_i^2 y_i = 1{,}419{,}780$, calcule SSE [$= \sum y_i^2 - \hat{\beta}_0 \sum y_i - \hat{\beta}_1 \sum x_i y_i - \hat{\beta}_2 \sum x_i^2 y_i$], y s.

d. Del inciso (c), SST $= 8386.43 - (210.70)^2/6 = 987.35$. Usando el SSE calculado en el inciso (c), ¿cuál es el valor calculado de R^2?

e. Si la desviación estándar estimada de $\hat{\beta}_2$ es $s_{\hat{\beta}_2} = .00226$, pruebe H_0: $\beta_2 = 0$ contra H_a: $\beta_2 \neq 0$ al nivel .01, e interprete el resultado.

29. En años recientes se han investigado de manera extensa los productos moldeables refractarios de alta alúmina por sus importantes ventajas sobre otros ladrillos refractarios de la misma clase, por ejemplo, menos costos de producción y aplicación, versatilidad y rendimiento a altas temperaturas. Los datos siguientes sobre x = viscosidad (MPa × s) y y = derrame libre (%) se obtuvieron de una gráfica del artículo "Processing of Zero-Cement Self-Flow Alumina Castables" (*The Amer. Ceramic Soc. Bull.*, 1998: 60-66):

x	351	367	373	400	402	456	484
y	81	83	79	75	70	43	22

Los autores del artículo científico citado relacionaron estas dos variables usando un modelo de regresión cuadrático. La función de regresión estimada es $y = -295.96 + 2.1885x - .0031662x^2$.

a. Calcule los valores y residuos pronosticados, y luego SSE y s^2.

b. Calcule e interprete el coeficiente de determinación múltiple.

c. La desviación estándar (SD) de $\hat{\beta}_2$ es $s_{\hat{\beta}_2} = .0004835$. ¿El predictor cuadrático pertenece al modelo de regresión?

d. La desviación estándar (SD) estimada de $\hat{\beta}_1$ es .4050. Use esto y la información en (c) para obtener los intervalos de confianza conjuntos para los coeficientes de regresión cuadrática y lineal con un nivel de confianza conjunto de (al menos) 95%.

e. La desviación estándar estimada de $\hat{\mu}_{Y\cdot 400}$ es 1.198. Calcule un intervalo de confianza de 95% para un derrame libre promedio real cuando la viscosidad = 400 y también un intervalo de predicción de 95% para derrame libre que resulte de

una sola observación hecha cuando la viscosidad es = 400, y compare los intervalos.

30. Los datos adjuntos fueron extraídos del artículo "Effects of Cold and Warm Temperatures on Springback of Aluminum-Magnesium Alloy 5083-H111" (*J. of Engr. Manuf.*, 2009: 427–431). La variable de respuesta es el límite elástico (MPa), y la predicción de la temperatura (°C).

x	−50	25	100	200	300
y	91.0	120.5	136.0	133.1	120.8

Aquí está la salida de Minitab del ajuste del modelo de regresión cuadrática (un gráfico en el documento citado sugiere que los autores lo hicieron):

```
Predictor          Coef    DE Coef       T       P
Constante       111.277      2.100   52.98   0.000
temperatura     0.32845    0.03303    9.94   0.010
temperatura
al cuadrado −0.0010050  0.0001213   −8.29   0.014

S = 3.44398 R-cuadrado = 98.1% R-cuadrado (adj) = 96.3%

Análisis de varianza

Fuente          GL        SS      MS      F      P
Regresión        2   1245.39  622.69  52.50  0.019
Error residual   2     23.72   11.86
Total            4   1269.11
```

a. ¿Qué proporción de la variación observada en la fuerza se puede atribuir a la relación del modelo?

b. Lleve a cabo una prueba de hipótesis al nivel de significancia .05 para decidir si el predictor cuadrático proporciona información útil por encima de la proporcionada por el predictor lineal.

c. Para un valor de fuerza de 100, $\hat{y} = 134.07$, $s_{\hat{Y}} = 2.38$. Estime la fuerza promedio real cuando la temperatura es de 100, de una manera que transmita información sobre la precisión y fiabilidad.

d. Utilice la información en (c) para predecir la fuerza de una única observación que se hizo cuando la temperatura es de 100 y hágalo de una manera que transmita información sobre la precisión y fiabilidad. Luego compare esta predicción con la estimación obtenida en (c).

31. Los datos adjuntos de y = producción de energía (W) y x = diferencia de temperatura (°K) fueron proporcionados por los autores del artículo "Comparison of Energy and Exergy Efficiency for Solar Box and Parabolic Cookers" (*J. of Energy Engr.*, 2007: 53–62).

Los autores del artículo ajustaron un modelo de regresión cúbico a los datos. Aquí está la salida de Minitab de tal ajuste.

```
La ecuación de regresión es

y = −134 + 12.7 x − 0.377 x**2 + 0.00359 x**3

Predictor          Coef    DE Coef       T       P
Constante      −133.787      8.048  −16.62   0.000
x               12.7423     0.7750   16.44   0.000
x**2           −0.37652    0.02444  −15.41   0.000
x**3          0.0035861  0.0002529   14.18   0.000

S = 0.168354 R-cuadrado = 98.0% R-cuadrado (adj) = 97.7%

Análisis de varianza

Fuente          GL        SS      MS      F      P
Regresión        3   27.9744  9.3248  329.00  0.000
Error residual  20    0.5669  0.0283
Total           23   28.5413
```

a. ¿Qué proporción de la variación observada en la producción de energía se puede atribuir a la relación del modelo?

b. Ajuste un modelo cuadrático para los datos resultantes en $R^2 = .780$. Calcular R^2 ajustado para este modelo y compárelo con R^2 ajustado para el modelo cúbico.

c. ¿El predictor cúbico parece proporcionar información útil sobre y por encima de la proporcionada por los predictores lineales y cuadráticos? Establezca y ponga a prueba las hipótesis adecuadas.

d. Cuando $x = 30$, $s_{\hat{Y}} = .0611$. Obtenga un IC del 95% de la producción real de energía promedio en este caso y también un 95% de IP para una producción de energía única que se observa cuando la diferencia de temperatura es de 30. *Sugerencia*: $s_{\hat{Y}} = .0611$.

e. Interprete las hipótesis H_0: $\mu_{Y \cdot 35} = 5$ frente a H_a: $\mu_{Y \cdot 35} \neq 5$ y luego lleve a cabo una prueba al nivel de significancia .05 con el hecho de que cuando $x = 35$, $s_{\hat{Y}} = .0523$.

32. La información siguiente es un subconjunto de datos obtenidos en un experimento para estudiar la relación entre el pH del suelo (x) y y = A1. La concentración/EC ("Root Responses of Three *Gramineae* Species to Soil Acidity in an Oxisol and an Ultisol", *Soil Science*, 1973: 295–302):

x	4.01	4.07	4.08	4.10	4.18
y	1.20	.78	.83	.98	.65
x	4.20	4.23	4.27	4.30	4.41
y	.76	.40	.45	.39	.30
x	4.45	4.50	4.58	4.68	4.70	4.77
y	.20	.24	.10	.13	.07	.04

Se propuso un modelo cúbico en el artículo, pero la versión de Minitab empleada por el autor del presente texto rechazó

x	23.20	23.50	23.52	24.30	25.10	26.20	27.40	28.10	29.30	30.60	31.50	32.01
y	3.78	4.12	4.24	5.35	5.87	6.02	6.12	6.41	6.62	6.43	6.13	5.92
x	32.63	33.23	33.62	34.18	35.43	35.62	36.16	36.23	36.89	37.90	39.10	41.66
y	5.64	5.45	5.21	4.98	4.65	4.50	4.34	4.03	3.92	3.65	3.02	2.89

incluir el término x^3 en el modelo, expresando que "x^3 está altamente correlacionada con otras variables predictoras". Para solucionar esto, $\bar{x} = 4.3456$ se restó de cada valor x para obtener $x' = x - \bar{x}$. Se requirió entonces una regresión cúbica para ajustar el modelo teniendo la función de regresión

$$y = \beta_0^* + \beta_1^* x' + \beta_2^* (x')^2 + \beta_3^* (x')^3$$

Dio por resultado la siguiente salida de computadora:

Parámetro	Estimado	DE estimada
β_0^*	.3463	.0366
β_1^*	−1.2933	.2535
β_2^*	2.3964	.5699
β_3^*	−2.3968	2.4590

a. ¿Cuál es la función de regresión estimada para el modelo "centrado"?
b. ¿Cuál es el valor estimado del coeficiente β_3 en el modelo "no centrado" con función de regresión $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$? ¿Cuál es la estimación de β_2?
c. Usando el modelo cúbico, ¿qué valor de y se pronosticaría cuando el pH del suelo es de 4.5?
d. Realice una prueba para determinar si el término cúbico debe ser retenido en el modelo.

33. En numerosos problemas de regresión con polinomios, en lugar de ajustar una función de regresión "centrada" usando $x' = x - \bar{x}$, la precisión en los cálculos se puede mejorar si se usa una función de la variable independiente estandarizada $x' = (x - \bar{x})/s_x$, donde s_x es la desviación estándar de las x_i. Considere ajustar la función de regresión cúbica $y = \beta_0^* + \beta_1^* x' + \beta_2^* (x')^2 + \beta_3^* (x')^3$ a los siguientes datos, que resultan de un estudio de la relación entre eficiencia de empuje y de cohetes impulsores supersónicos y el ángulo x de semidivergencia de la nariz del cohete ("More on Correlating Data", *CHEMTECH*, 1976: 266–270):

x	5	10	15	20	25	30	35
y	.985	.996	.988	.962	.940	.915	.878

Parámetro	Estimado	DE estimada
β_0^*	.9671	.0026
β_1^*	−.0502	.0051
β_2^*	−.0176	.0023
β_3^*	.0062	.0031

a. ¿Qué valor de y se pronosticaría cuando el ángulo de semidivergencia sea 20? ¿Y cuando $x = 25$?
b. ¿Cuál es la función de regresión estimada $\hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2 + \hat{\beta}_3 x^3$ para el modelo "no estandarizado"?
c. Use una prueba de nivel .05 para determinar si el término cúbico debe borrarse del modelo.
d. ¿Qué se puede decir acerca de la relación entre las SSE y las R^2 para modelos estandarizados y no estandarizados? Explique.
e. La SSE para el modelo cúbico es .00006300, mientras que para un modelo cuadrático la SSE es .00014367. Calcule la R^2 para cada modelo. ¿La diferencia entre las dos sugiere que el término cúbico debe ser borrado?

34. La información siguiente resultó de un experimento para evaluar el potencial de tierras sin quemar de una mina de carbón, como medio para el crecimiento de plantas. Las variables son x = cationes extraíbles de ácido y y = acidez intercambiable/capacidad total de intercambio de cationes ("Exchangeable Acidity in Unburnt Colliery Spoil", *Nature*, 1969: 161):

x	−23	−5	16	26	30	38	52
y	1.50	1.46	1.32	1.17	.96	.78	.77
x	58	67	81	96	100	113	
y	.91	.78	.69	.52	.48	.55	

La estandarización de la variable independiente x para obtener $x' = (x - \bar{x})/s_x$, y el ajuste de la función de regresión $y = \beta_0^* + \beta_1^* x' + \beta_2^* (x')^2$, dieron la siguiente salida de computadora.

Parámetro	Estimado	DE estimada
β_0^*	.8733	.0421
β_1^*	−.3255	.0316
β_2^*	.0448	.0319

a. Estime $\mu_{Y \cdot 50}$.
b. Calcule el valor del coeficiente de determinación múltiple (vea el ejercicio 28(c)).
c. ¿Cuál es la función de regresión estimada $\hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2$ usando la variable x no estandarizada?
d. ¿Cuál es la desviación estándar estimada de $\hat{\beta}_2$ calculada en el inciso (c)?
e. Realice una prueba usando las estimaciones estandarizadas para determinar si el término cuadrático debe retenerse en el modelo. Repita usando las estimaciones no estandarizadas. ¿Difieren sus conclusiones?

35. El artículo "The Respiration in Air and in Water of the Limpets *Patella caerulea* and *Patella lusitanica*" (*Comp. Biochemistry and Physiology*, 1975: 407-411) propuso un sencillo modelo de potencia para la relación entre el ritmo de respiración y y la temperatura x para *P. caerulea* en aire. No obstante, una gráfica de $\ln(y)$ en función de x muestra una figura curva. Ajuste el modelo cuadrático de potencia $Y = \alpha e^{\beta x + \gamma x^2} \cdot \epsilon$ a los datos siguientes.

x	10	15	20	25	30
y	37.1	70.1	109.7	177.2	222.6

13.4 Análisis de regresión múltiple

En regresión múltiple, el objetivo es construir un modelo probabilístico que relacione una variable dependiente y a más de una variable independiente o predictor. Represente con k el número de variables predictoras ($k \geq 2$) y denote estas predictores por $x_1, x_2, \ldots, x_k$. Por ejemplo, al tratar de predecir el precio de venta de una casa, se podría tener $k = 3$ con $x_1 =$ tamaño (ft^2), $x_2 =$ edad (años), y $x_3 =$ número de habitaciones.

DEFINICIÓN

La **ecuación general del modelo de regresión múltiple aditivo** es

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon \qquad (13.15)$$

donde $E(\epsilon) = 0$ y $V(\epsilon) = \sigma^2$. Además, para fines de prueba de hipótesis y calcular intervalos de confianza o intervalos de predicción, se supone que ϵ está normalmente distribuida.

Sean $x_1^*, x_2^*, \ldots, x_k^*$ valores particulares de $x_1, \ldots, x_k$. Entonces (13.15) implica que

$$\mu_{Y \cdot x_1^*, \ldots, x_k^*} = \beta_0 + \beta_1 x_1^* + \cdots + \beta_k x_k^* \qquad (13.16)$$

En esta forma, así como $\beta_0 + \beta_1 x$ describe el valor medio Y como función de x en regresión lineal simple, la **verdadera función de regresión** (o **población**) $\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$ da el valor esperado de Y como función de $x_1, \ldots, x_k$. Las β_i son los **verdaderos coeficientes de regresión** (o **población**). El coeficiente de regresión β_1 se interpreta como el cambio esperado en Y asociado con un aumento de una unidad en x_1 *mientras* $x_2, \ldots, x_k$ *se mantengan fijos*. Interpretaciones análogas se cumplen para $\beta_2, \ldots, \beta_k$.

Modelos con interacción y predictores cuadráticos

Si un investigador ha obtenido observaciones en y, x_1 y x_2, un modelo posible es $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$. No obstante, se pueden construir otros modelos al formar predictores y funciones matemáticas de x_1 y/o x_2. Por ejemplo, con $x_3 = x_1^2$ y $x_4 = x_1 x_2$, el modelo

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \epsilon$$

tiene la forma general de (13.15). En general, no es sólo permisible para algunos predictores ser funciones matemáticas de otras sino también que, con frecuencia, sean altamente deseables en el sentido de que el modelo resultante pueda ser mucho más exitoso para explicar la variación en y que cualquier otro modelo sin estos predictores. Esta discusión también muestra que la regresión con polinomios es ciertamente un caso especial de regresión múltiple. Por ejemplo, el modelo cuadrático $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon$ tiene la forma de (13.15) con $k = 2$, $x_1 = x$ y $x_2 = x^2$.

Para el caso de dos variables independientes, x_1 y x_2, hay cuatro modelos útiles de regresión múltiple.

1. El modelo de primer orden:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

2. El modelo de segundo orden sin interacción:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \epsilon$$

3. El modelo con predictores de primer orden e interacción:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon$$

4. El modelo de segundo orden completo o cuadrático completo:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \epsilon$$

La comprensión de las diferencias entre estos modelos es un primer paso importante en la construcción de modelos de regresión realistas a partir de las variables independientes bajo estudio.

El modelo de primer orden es la generalización más fácil de regresión lineal simple. Expresa que para un valor fijo de cualquiera de las dos variables, el valor esperado de Y es una función lineal de la otra variable y que el cambio esperado en Y asociado con un aumento unitario en x_1 (x_2) es β_1 (β_2) independiente del nivel de x_2 (x_1). Entonces, si se grafica la función de regresión como una función de x_1 para diversos valores diferentes de x_2, se obtienen como contornos de la función de regresión un conjunto de rectas paralelas, como se ve en la figura 13.13(a). La función $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ especifica un plano en espacio tridimensional; el primer modelo dice que cada uno de los valores observados de la variable dependiente corresponde a un punto que se desvía verticalmente de este plano en una cantidad aleatoria ϵ.

Según el modelo de segundo orden sin interacción, si x_2 es fija, el cambio esperado en Y para un aumento de 1 unidad en x_1 es

$$\beta_0 + \beta_1(x_1 + 1) + \beta_2 x_2 + \beta_3(x_1 + 1)^2 + \beta_4 x_2^2$$

$$- (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2) = \beta_1 + \beta_3 + 2\beta_3 x_1$$

Debido a que este cambio esperado no depende de x_2, los contornos de la función de regresión para diferentes valores de x_2 son todavía paralelos entre sí. No obstante, la dependencia del cambio esperado en el valor de x_1 significa que los contornos son ahora curvas en lugar de rectas. Esto se ve en la figura 13.13(b). En este caso, la superficie de regresión ya no es un plano en espacio tridimensional sino que es una superficie curvada.

Los contornos de la función de regresión para el modelo de primer orden con interacción son rectas no paralelas. Esto es porque el cambio esperado en Y cuando x_1 se aumenta en 1 es

$$\beta_0 + \beta_1(x_1 + 1) + \beta_2 x_2 + \beta_3(x_1 + 1)x_2$$

$$- (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2) = \beta_1 + \beta_3 x_2$$

Este cambio esperado depende del valor de x_2, de modo que cada línea de contorno debe tener una pendiente diferente, como se ve en la figura 13.13(c). La palabra *interacción* refleja el hecho de que un cambio esperado en Y, cuando una variable aumenta en valor, depende del valor de la otra variable.

Por último, para el modelo completo de segundo orden, el cambio esperado en Y cuando x_2 se mantiene fijo mientras x_1 aumenta en 1 unidad es $\beta_1 + \beta_3 + 2\beta_3 x_1 + \beta_5 x_2$, que es una función de x_1 y de x_2. Esto implica que los contornos de la función de regresión son curvados y no paralelos entre sí, como se ilustra en la figura 13.13(d).

Similares consideraciones aplican a modelos construidos a partir de más de dos variables independientes. En general, la presencia de términos de interacción en el modelo implica que el cambio esperado en Y depende no sólo de la variable que se aumenta o disminuye sino también de los valores de algunas de las variables fijas. Al igual que en ANOVA, es posible tener términos de interacción de avance más elevado (por ejemplo $x_1 x_2 x_3$), lo que hace más difícil la interpretación del modelo.

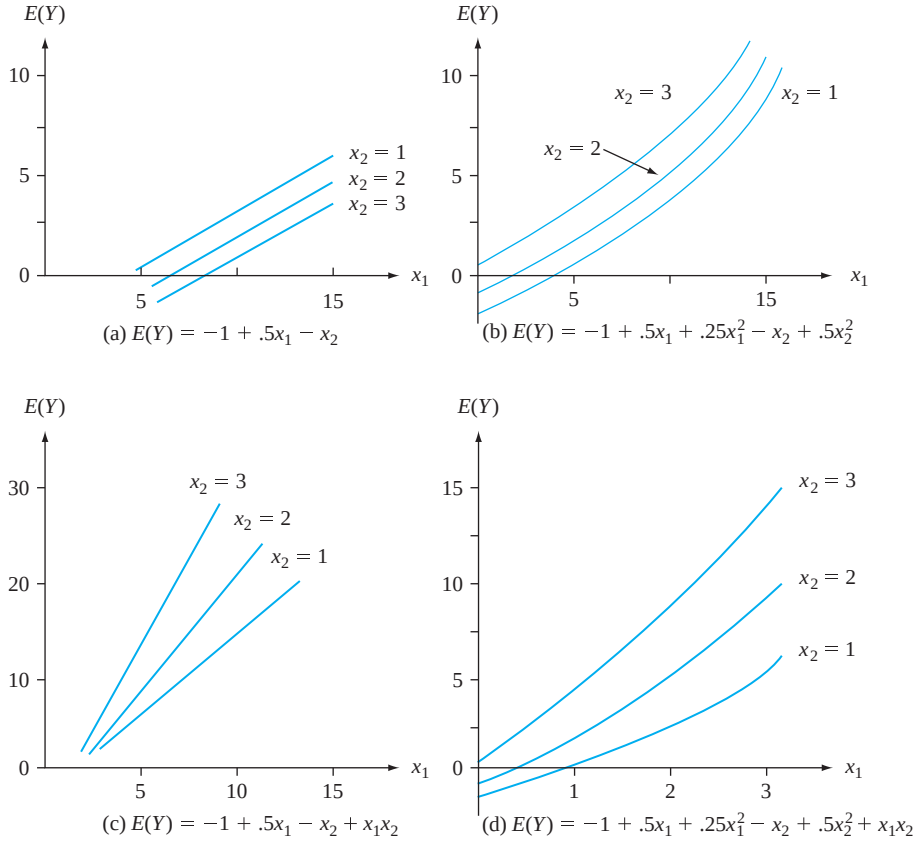


Figura 13.13 Contornos de cuatro funciones de regresión diferentes

Nótese que si el modelo contiene predictores de interacción o cuadráticos, la interpretación genérica de una β_i dada previamente no es aplicable por lo general. Esto es porque entonces no es posible aumentar x_i en 1 unidad y mantener fijos los valores de todos los otros predictores.

Modelos con predictores para variables categóricas

Hasta este punto se ha considerado explícitamente la inclusión de sólo variables predictoras cuantitativas (numéricas) en un modelo de regresión múltiple. Con el uso de codificación numérica sencilla, las variables cualitativas (categóricas), por ejemplo, material para cojinetes (aluminio o cobre/plomo) o tipo de madera (pino, roble, o nogal), también se pueden incorporar en un modelo. Hay que enfocarse primero en el caso de una variable dicotómica, una con sólo dos categorías posibles, hombre o mujer, de manufactura norteamericana o extranjera, etcétera. Con cualquiera de estas variables, se asocia una **variable indicadora *x*** o **imaginaria** cuyos posibles valores 0 y 1 indican qué categoría es relevante para cualquier observación particular.

Ejemplo 13.11 El artículo "Estimating Urban Travel Times: A Comparative Study" (*Trans. Res.*, 1980: 173–175) describió un estudio que relacionaba la variable dependiente y = tiempo de viaje entre lugares en cierta ciudad y la variable independiente x_2 = distancia entre lugares. Dos tipos de vehículos, autos de pasajeros y camiones, se emplearon en el estudio. Sea

$$x_1 = \begin{cases} 1 & \text{si el vehículo es un camión} \\ 0 & \text{si el vehículo es un auto de pasajeros} \end{cases}$$

Un posible modelo de regresión múltiple es

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

El valor medio de tiempo de viaje depende de si un vehículo es un auto o un camión:

$$\text{tiempo medio} = \beta_0 + \beta_2 x_2 \qquad \text{cuando } x_1 = 0 \text{ (autos)}$$

$$\text{tiempo medio} = \beta_0 + \beta_1 + \beta_2 x_2 \qquad \text{cuando } x_1 = 1 \text{ (camiones)}$$

El coeficiente β_1 es la diferencia en tiempos medios entre camiones y autos con la distancia mantenida fija; si $\beta_1 > 0$, a los camiones les tomará más tiempo en promedio recorrer cualquier distancia particular que a los autos.

Una segunda posibilidad es un modelo con un predictor de interacción:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon$$

Ahora los tiempos medios para los dos tipos de vehículos son

$$\text{tiempo medio} = \beta_0 + \beta_2 x_2 \qquad \text{cuando } x_1 = 0$$

$$\text{tiempo medio} = \beta_0 + \beta_1 + (\beta_2 + \beta_3) x_2 \qquad \text{cuando } x_1 = 1$$

Para cada modelo, la gráfica del tiempo medio contra distancia es una recta para cualquiera de los dos tipos de vehículo, como se ilustra en la figura 13.14. Las dos rectas son paralelas para el primer modelo (sin interacción), pero en general tendrán diferentes pendientes cuando el segundo modelo es correcto. Para este último modelo, el cambio en tiempo medio de viaje asociado con un aumento de 1 milla en distancia depende de qué tipo de vehículo se trata, las dos variables "tipo de vehículo" y "tiempo de viaje" interactúan. De hecho, los datos recolectados por los autores del artículo citado líneas antes sugirieron la presencia de interacción.

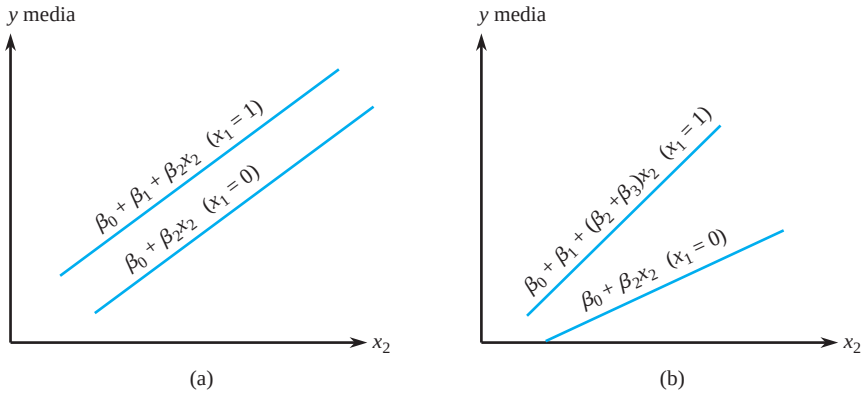


Figura 13.14 Funciones de regresión para modelos con una variable imaginaria (x_1) y una variable cuantitativa x_2: (a) sin interacción; (b) interacción

Se podría pensar que la forma de manejar una situación de tres categorías es definir una sola variable numérica con valores codificados, como por ejemplo 0, 1 y 2 correspondientes a las tres categorías. Esto es incorrecto, porque impone un orden en las categorías que no está necesariamente implicado por el contexto del problema. El método correcto para incorporar tres categorías es definir *dos* variables imaginarias diferentes. Suponga, por ejemplo, que y es la vida útil de cierta herramienta de corte, x_1 es la velocidad de corte, y que hay tres marcas de herramienta que se investigan. Entonces, sea

$$x_2 = \begin{cases} 1 & \text{si se usa la herramienta marca A} \\ 0 & \text{de otro modo} \end{cases} \qquad x_3 = \begin{cases} 1 & \text{si se usa la herramienta marca B} \\ 0 & \text{de otro modo} \end{cases}$$

Cuando se hace una observación en una herramienta marca A, $x_2 = 1$ y $x_3 = 0$, mientras que para una herramienta marca B, $x_2 = 0$ y $x_3 = 1$. Una observación hecha en una herramienta marca C tiene $x_2 = x_3 = 0$, y no es posible que $x_2 = x_3 = 1$ porque una herramienta no puede ser al mismo tiempo marca A y marca B. El modelo sin interacción tendría sólo los predictores x_1, x_2 y x_3. El siguiente modelo con interacción permite que el cambio medio en duración, asociado con un aumento de 1 unidad en velocidad, dependa de la marca de herramienta:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \epsilon$$

La construcción de una imagen como la figura 13.14, con una gráfica para cada uno de los tres pares posibles (x_2, x_3), da tres líneas no paralelas (a menos que $\beta_4 = \beta_5 = 0$).

En forma más general, incorporar una variable categórica con c posibles categorías en un modelo de regresión múltiple requiere el uso de $c - 1$ variables indicadoras (por ejemplo, cinco marcas de herramientas necesitarían usar cuatro variables indicadoras). Entonces, incluso una variable categórica puede sumar numerosos predictores a un modelo.

Estimación de parámetros

Los datos en regresión lineal simple constan de n pares $(x_1, y_1), \ldots, (x_n, y_n)$. Suponga que un modelo de regresión múltiple contiene dos variables predictoras, x_1 y x_2. Entonces, el conjunto de datos estará formado por n ternas $(x_{11}, x_{21}, y_1), (x_{12}, x_{22}, y_2), \ldots, (x_{1n}, x_{2n}, y_n)$. Aquí el primer subíndice de x se refiere al predictor y el segundo al número de observación. Más generalmente, con k predictores, los datos constan de $n(k + 1)$ tuplas $(x_{11}, x_{21}, \ldots, x_{k1}, y_1), (x_{12}, x_{22}, \ldots, x_{k2}, y_2), \ldots, (x_{1n}, x_{2n}, \ldots, x_{kn}, y_n)$, donde x_{ij} es el valor del i-ésimo predictor x_i asociado con el valor observado y_j. Se supone que las y_j han sido observadas independientemente entre sí de acuerdo con el modelo (13.15). Para estimar los parámetros $\beta_0, \beta_1, \ldots, \beta_k$ usando el principio de mínimos cuadrados, forme la suma de desviaciones cuadradas de las y_j observadas desde una función de ensayo $y = b_0 + b_1 x_1 + \cdots + b_k x_k$:

$$f(b_0, b_1, \ldots, b_k) = \sum_j [y_j - (b_0 + b_1 x_{1j} + b_2 x_{2j} + \cdots + b_k x_{kj})]^2 \qquad (13.17)$$

Las estimaciones de mínimos cuadrados son los valores de las b_i que minimizan $f(b_0, \ldots, b_k)$. Si se toma la derivada parcial de f con respecto a cada una de las b_i $(i = 0, 1, \ldots, k)$ y se igualan a cero todas las parciales, se obtiene el siguiente sistema de **ecuaciones normales:**

$$\begin{aligned}
&b_0 n + b_1 \textstyle\sum x_{1j} + b_2 \sum x_{2j} + \cdots + b_k \sum x_{kj} = \sum y_j \\
&b_0 \textstyle\sum x_{1j} + b_1 \sum x_{1j}^2 + b_2 \sum x_{1j} x_{2j} + \cdots + b_k \sum x_{1j} x_{kj} = \sum x_{1j} y_j \\
&\quad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\
&b_0 \textstyle\sum x_{kj} + b_1 \sum x_{1j} x_{kj} + \cdots + b_{k-1} \sum x_{k-1,j} x_{kj} + b_k \sum x_{kj}^2 = \sum x_{kj} y_j
\end{aligned} \qquad (13.18)$$

Estas ecuaciones son lineales en las incógnitas $b_0, b_1, \ldots, b_k$. Al resolver (13.18) se obtienen las estimaciones de mínimos cuadrados $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$. Esto se hace mejor si se utiliza un paquete de software de estadística.

Ejemplo 13.12 El artículo "How to Optimize and Control the Wire Bonding Process: Part II" (*Solid State Technology*, enero de 1991: 67–72) describió un experimento realizado para evaluar el impacto de las variables x_1 = fuerza (gm), x_2 = potencia (mW), x_3 = temperatura (°C), y x_4 = tiempo (ms) en y = resistencia al corte (gm). Los datos siguientes* se generaron para ser consistentes con la información dada en el artículo:

Observación	Fuerza	Potencia	Temperatura	Tiempo	Resistencia
1	30	60	175	15	26.2
2	40	60	175	15	26.3
3	30	90	175	15	39.8
4	40	90	175	15	39.7
5	30	60	225	15	38.6
6	40	60	225	15	35.5
7	30	90	225	15	48.8
8	40	90	225	15	37.8
9	30	60	175	25	26.6
10	40	60	175	25	23.4
11	30	90	175	25	38.6
12	40	90	175	25	52.1
13	30	60	225	25	39.5
14	40	60	225	25	32.3
15	30	90	225	25	43.0
16	40	90	225	25	56.0
17	25	75	200	20	35.2
18	45	75	200	20	46.9
19	35	45	200	20	22.7
20	35	105	200	20	58.7
21	35	75	150	20	34.5
22	35	75	250	20	44.0
23	35	75	200	10	35.7
24	35	75	200	30	41.8
25	35	75	200	20	36.5
26	35	75	200	20	37.6
27	35	75	200	20	40.3
28	35	75	200	20	46.0
29	35	75	200	20	27.8
30	35	75	200	20	40.3

Un paquete computarizado de estadística dio los siguientes estimados de mínimos cuadrados:

$$\hat{\beta}_0 = -37.48 \quad \hat{\beta}_1 = .2117 \quad \hat{\beta}_2 = .4983 \quad \hat{\beta}_3 = .1297 \quad \hat{\beta}_4 = .2583$$

Entonces se estima que .1297 gm es el cambio promedio en resistencia asociado con un aumento de 1 grado en temperatura, cuando los otros tres predictores se mantienen fijos; los otros coeficientes estimados se interpretan de un modo semejante.

La ecuación estimada de regresión es

$$y = -37.48 + .2117x_1 + .4983x_2 + .1297x_3 + .2583x_4$$

Una predicción puntual de resistencia que resulta de una fuerza de 35 gm, potencia de 75 mW, temperatura de 200° y tiempo de 20 ms es

$$\begin{aligned}\hat{y} &= -37.48 + (.2117)(35) + (.4983)(75) + (.1297)(200) + (.2583)(20) \\ &= 38.41\,\text{gm}\end{aligned}$$

* Del libro *Statistics Engineering Problem Solving*, de Stephen Vardeman, una excelente exposición del área cubierta en este libro, aunque a un nivel un tanto más alto.

Ésta también es una estimación puntual del valor medio de resistencia para los valores especificados de fuerza, potencia, temperatura y tiempo. ■

R^2 y $\hat{\sigma}^2$

Predecir o ajustar valores, residuos y las diversas sumas de cuadrados se calculan como en la regresión lineal simple y con polinomio. El valor predicho de $\hat{y}_1$ resulta de la sustitución de los valores de los diferentes predictores desde la primera observación en la función de regresión estimada:

$$\hat{y}_1 = \hat{\beta}_0 + \hat{\beta}_1 x_{11} + \hat{\beta}_1 x_{21} + \cdots + \hat{\beta}_k x_{k1}$$

Los valores restantes predichos $\hat{y}_2, \ldots, \hat{y}_n$ proceden de la sustitución de los valores de los predictores de la 2ª, 3ª, . . . , y finalmente la n-ésima observación en la función estimada. Por ejemplo, los valores de los cuatro predictores de la última observación en el ejemplo 13.12 son $x_{1,30} = 35$, $x_{2,30} = 75$, $x_{3,30} = 200$, y $x_{4,30} = 20$ por tanto

$$\hat{y}_{30} = -37.48 + .2117(35) + .4983(75) + .1297(200) + .2583(20) = 38.41$$

Los residuos $y_1 - \hat{y}_1, \ldots, y_n - \hat{y}_n$ son las diferencias entre los valores observados y los predichos. El último residuo del ejemplo 13.12 es $40.3 - 38.41 = 1.89$. Cuanto más cerca de 0 se encuentren los residuos, mejor será el trabajo que haga la función de regresión estimada para predecir los valores correspondientes a las observaciones en la muestra.

La suma residual de error o de cuadrados es $\text{SSE} = \sum(y_i - \hat{y}_i)^2$. De nuevo esto es interpretado como una medida de cómo una gran variación en los valores observados de y no se explica por (no se atribuye a) la relación del modelo. El número de gl asociados con SSE es $n - (k + 1)$, ya que $k + 1$ gl se pierden en la estimación de $k + 1$ coeficientes β. La suma total de cuadrados, una medida de la variación total en los valores y observados, es $\text{SST} = \sum(y_i - \bar{y})^2$. La suma de regresión de cuadrados $\text{SSR} = \sum(\hat{y}_i - \bar{y})^2 = \text{SST} - \text{SSE}$ es una medida de la variación explicada. Entonces el **coeficiente de determinación múltiple R^2** es

$$R^2 = 1 - \text{SSE/SST} = \text{SSR/SST}$$

Esto se interpreta como la proporción de la variación observada y que puede ser explicada por el ajuste a los datos por el modelo de regresión múltiple.

Debido a que no existe una visión preliminar de los datos de regresión múltiple similar a un gráfico de dispersión para los datos bivariados, el coeficiente de determinación múltiple es nuestra primera indicación de si el modelo elegido tiene éxito en la explicación de la variación de y. Por desgracia, hay un problema con R^2: su valor puede ser inflado por la adición de un gran número de factores predictivos en el modelo, incluso si la mayoría de estos predictores son más bien frívolos. Por ejemplo, supongamos que y es el precio de venta de una casa. Luego predictores sensibles incluyen x_1 = el tamaño interior de la casa, x_2 = el tamaño del terreno en el que la casa se asienta, x_3 = el número de dormitorios, x_4 = el número de cuartos de baño y x_5 = edad de la casa. Ahora supongamos que añadimos x_6 = el diámetro de la manija de la puerta en el armario de los abrigos, x_7 = espesor de la tabla de picar en la cocina, x_8 = espesor de la losa de patio y así sucesivamente. A menos que tengamos muy mala suerte en nuestra elección de los predictores, utilizando $n - 1$ predictores (uno menos que el tamaño de la muestra) resultará $R^2 = 1$. Así que el objetivo en la regresión múltiple no es simplemente para explicar la mayor parte de la variación observada de y, pero para hacerlo utilizamos un modelo con relativamente pocos predictores que son fáciles de interpretar. Por tanto, es deseable ajustar R^2, como se hizo en la regresión polinómica, para tener en cuenta el tamaño del modelo:

$$R_a^2 = 1 - \frac{\text{SSE}/(n - (k + 1))}{\text{SST}/(n - 1)} = 1 - \frac{n - 1}{n - (k + 1)} \cdot \frac{\text{SSE}}{\text{SST}}$$

Debido a la razón enfrente de la SSE/SST excede a 1, R_a^2 es menor que R^2. Además, cuanto mayor sea el número k de predictores en relación con el tamaño de muestra n, R_a^2 menor será relativa a R^2. Este R^2 ajustado, incluso puede ser negativo, mientras que R^2 debe estar entre 0 y 1. Un valor R_a^2 que es sustancialmente más pequeño que R^2 es en sí una advertencia de que el modelo puede contener muchos predictores.

La raíz cuadrada positiva de R^2 se llama el *coeficiente de correlación múltiple* y se denota por R. Se puede demostrar que R es el coeficiente de correlación muestral calculado a partir de los pares $(\hat{y}_i, y_i)$ (es decir, utilizar $\hat{y}_i$ en lugar de x_i en la fórmula para r de la sección 12.5).

SSE es también la base para estimar los parámetros restantes del modelo:

$$\hat{\sigma}^2 = s^2 = \frac{\text{SSE}}{n - (k + 1)} = \text{MSE}$$

Ejemplo 13.13 Unos investigadores llevaron a cabo un estudio para ver la forma en que diversas características del concreto se ven afectadas por x_1 = % de piedra caliza en polvo y x_2 = proporción de agua y cemento; dicho estudio dio por resultado los datos que aparecen a continuación ("Durability of Concrete with Addition of Limestone Powder", *Magazine of Concrete Research*, 1996; 131–137).

x_1	x_2	x_1x_2	Resist. a la compr. de 28 días (MPa)	Adsorción (%)
21	.65	13.65	33.55	8.42
21	.55	11.55	47.55	6.26
7	.65	4.55	35.00	6.74
7	.55	3.85	35.90	6.59
28	.60	16.80	40.90	7.28
0	.60	0.00	39.10	6.90
14	.70	9.80	31.55	10.80
14	.50	7.00	48.00	5.63
14	.60	8.40	42.30	7.43
			$\bar{y} = 39.317$, SST = 278.52	$\bar{y} = 7.339$, SST = 18.356

Considere primero la resistencia a la compresión como la variable dependiente y. El ajuste del modelo de primer orden da por resultado

$$y = 84.82 + .1643x_1 - 79.67x_2,\ \text{SSE} = 72.52\ (\text{gl} = 6),\ R^2 = .741,\ R_a^2 = .654$$

mientras que si se incluye un predictor de interacción dará

$$y = 6.22 + 5.779x_1 + 51.33x_2 - 9.357x_1x_2$$
$$\text{SSE} = 29.35\ (\text{gl} = 5) \qquad R^2 = .895 \qquad R_a^2 = .831$$

Con base en este último ajuste, una predicción para la resistencia a la compresión cuando el porcentaje de piedra caliza es 14 y la proporción de agua y cemento = .60 es

$$\hat{y} = 6.22 + 5.779(14) + 51.33(.60) - 9.357(8.4) = 39.32$$

Un ajuste de toda la relación cuadrática da por resultado que prácticamente no hay cambio en el valor de R^2. No obstante, cuando la variable dependiente es la adsorción, se obtienen los siguientes resultados: $R^2 = .747$ cuando se usan sólo dos predictores, .802 cuando se agrega el predictor de interacción y .889 cuando se usan los cinco predictores para toda la relación cuadrática. ■

En general, $\hat{\beta}_i$ se puede interpretar como una estimación del cambio promedio en Y asociado con un aumento de 1 unidad en x_i mientras se mantengan fijos los valores de

todas las otros pronosticadores. A veces, sin embargo, es difícil y hasta imposible aumentar el valor de un predictor al tiempo que se mantienen fijos todos los otros. En situaciones como éstas, hay una interpretación alternativa de los coeficientes de regresión estimados. En síntesis, supóngase que $k = 2$, y se denota con $\hat{\beta}_1$ la estimación de β_1 en la regresión de y en los dos pronosticadores x_1 y x_2. Entonces

1. Haga regresión de y contra sólo x_2 (una regresión lineal simple) y denote con $g_1, g_2, \ldots, g_n$ los residuos resultantes. Estos residuos representan variación en y después de eliminar o ajustar los efectos de x_2.
2. Haga regresión de x_1 contra x_2 (esto es, considere x_1 como la variable dependiente y x_2 la variable independiente en esta regresión lineal simple), y denote los residuos por $f_1, \ldots, f_n$. Estos residuos representan variación en x_1 después de eliminar o ajustar los efectos de x_2.

Ahora considere graficar los residuos de la primera regresión contra los de la segunda; esto es, grafique los pares $(f_1, g_1), \ldots, (f_n, g_n)$. El resultado se denomina *gráfica parcial de residuos* o *gráfica ajustada de residuos.* Si ha de ajustarse una recta de regresión a los puntos de esta gráfica, la pendiente resulta ser exactamente $\hat{\beta}_1$ (además, los residuos de esta recta son exactamente los residuos $e_1, \ldots, e_n$ de la regresión múltiple de y en x_1 y x_2). Entonces, $\hat{\beta}_1$ se puede interpretar como los cambios estimados en y asociados con un aumento de 1 unidad en x_1 después de eliminar o ajustar los efectos de cualesquiera otros pronosticadores del modelo. La misma interpretación se cumple para otros coeficientes estimados, cualquiera que sea el número de predictores del modelo (no hay algo especial acerca de $k = 2$; el argumento anterior sigue válido si se hace regresión de y contra todos los pronosticadores que no sean x_1 en el paso 1 y se hace regresión de x_1 contra los otros $k - 1$ pronosticadores del paso 2).

Como ejemplo, suponga que y es el precio de venta de un edificio de departamentos y que los predictores son números de departamento, antigüedad, tamaño de lote, número de espacios de estacionamiento, y área total del edificio (pie^2). Puede no ser razonable aumentar el número de departamentos sin incrementar también el área total. No obstante, si $\hat{\beta}_5 = 16.00$, entonces se puede decir que un aumento de \$16 en el precio de venta está asociado con cada pie cuadrado extra de área total, después de ajustar los efectos de los otros cuatro predictores.

Una prueba de utilidad de modelo

Con datos multivariados, no hay una representación preliminar análoga a un diagrama de dispersión para indicar si un modelo particular de regresión múltiple explica exitosamente la variación observada de y. El valor de R^2 ciertamente comunica un mensaje preliminar, pero este valor es a veces engañoso porque puede estar fuertemente inflado si se usa un número grande de predictores con respecto al tamaño muestral. Por esta razón, es importante tener una prueba formal para la utilidad del modelo.

La prueba de la utilidad de un modelo en regresión lineal simple comprendió la hipótesis nula H_0: $\beta_1 = 0$, según la cual no hay relación útil entre y y el predictor individual x. Aquí se considera la afirmación de que $\beta_1 = 0, \beta_2 = 0, \ldots, \beta_k = 0$, que dice que no hay relación útil entre y y *cualquiera* de los k predictores. Si al menos una de estas β no es 0, el (los) predictor(es) correspondiente(s) es(son) útil(es). La prueba está basada en un estadístico que tiene una distribución F particular cuando H_0 es verdadera.

Hipótesis nula: H_0: $\beta_1 = \beta_2 = \cdots = \beta_k = 0$

Hipótesis alternativa: H_a: al menos una $\beta_i \neq 0$ $\quad (i = 1, \ldots, k)$

Valor del estadístico de prueba: $f = \dfrac{R^2/k}{(1 - R^2)/[n - (k + 1)]}$

$$= \frac{\text{SSR}/k}{\text{SSE}/[n-(k+1)]} = \frac{\text{MSR}}{\text{MSE}} \qquad (13.19)$$

donde SSR = suma de cuadrados de regresión = SST − SSE

Región de rechazo para una prueba de nivel α: $f \geq F_{\alpha,k,n-(k+1)}$

Excepto para un múltiplo constante, el estadístico de prueba aquí es $R^2/(1-R^2)$, que es la razón entre una variación explicada a una no explicada. Si la proporción de variación explicada es alta con respecto a la no explicada, naturalmente se rechazaría H_0 y se confirmaría la utilidad del modelo. No obstante, si k es grande con respecto a n, el factor $[(n-(k+1))/k]$ reducirá considerablemente a f.

Ejemplo 13.14 Volviendo a los datos de resistencia del pegamento al corte del ejemplo 13.12, se ajustó un modelo con $k = 4$ predictores, de manera que las hipótesis relevantes son

H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

H_a: al menos una de estas cuatro β no es 0

La figura 13.15 muestra la salida impresa del paquete de estadística JMP. Los valores de s (raíz cuadrada del error de la media), R^2 y R^2 ajustada ciertamente sugieren un modelo útil. El valor de la razón F de utilidad del modelo es

$$f = \frac{R^2/k}{(1-R^2)/[n-(k+1)]} = \frac{.713959/4}{.286041/(30-5)} = 15.60$$

Response: strength

Summary of Fit

RSquare	0.713959
RSquare Adj	0.668193
Root Mean Square Error	5.157979
Mean of Response	38.40667
Observations (or Sum Wgts)	30

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-37.47667	13.09964	-2.86	0.0084
force	0.2116667	0.210574	1.01	0.3244
power	0.4983333	0.070191	7.10	<.0001
temp	0.1296667	0.042115	3.08	0.0050
time	0.2583333	0.210574	1.23	0.2313

Whole-Model Test

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	1660.1400	415.035	15.6000
Error	25	665.1187	26.605	Prob>F
C Total	29	2325.2587		<.0001

Figura 13.15 Salida de regresión múltiple del JMP para los datos del ejemplo 13.14

Este valor también aparece en la columna F Ratio de la tabla ANOVA de la figura 13.15. El máximo valor crítico F para grado de libertad de numerador 4 y denominador 25 en la tabla A.9 del apéndice es 6.49, que captura un área de cola superior de .001. De aquí el valor $P < .001$. La tabla ANOVA de la salida impresa del JMP muestra que el valor $P < .0001$. Éste es un resultado muy significativo. La hipótesis nula debería ser rechazada a cualquier nivel razonable de significación. La conclusión es que *hay* una relación lineal útil entre y o *al menos una* de los cuatro predictores del modelo. Esto no significa que los cuatro predictores sean útiles; un poco más adelante se tratará más de esto. ■

Inferencias en regresión múltiple

Antes de construir hipótesis y los IC y hacer predicciones, la adecuación del modelo debe ser evaluada e investigado el impacto de las observaciones atípicas. Los métodos para hacer esto se describen al final de la presente sección y en la siguiente.

Debido a que cada $\hat{\beta}_i$ es una función lineal de las y_i, la desviación estándar de cada $\hat{\beta}_i$ es el producto de σ y una función de las x_{ij}, de modo que se obtiene una estimación $s_{\hat{\beta}_i}$ para esta desviación estándar al sustituir s con σ. La función de las x_{ij} es bastante complicada, pero todos los paquetes computarizados estándares de estadística calculan y muestran las $s_{\hat{\beta}_i}$. Las inferencias respecto a una sola β_i están basadas en la variable estandarizada

$$T = \frac{\hat{\beta}_i - \beta_i}{S_{\hat{\beta}_i}}$$

que tiene una distribución t con $n - (k + 1)$ grados de libertad.

La estimación puntual de $\mu_{Y \cdot x_1^*, \ldots, x_k^*}$, el valor esperado de Y cuando $x_1 = x_1^*, \ldots, x_k = x_k^*$, es $\hat{\mu}_{Y \cdot x_1^*, \ldots, x_k^*} = \hat{\beta}_0 + \hat{\beta}_1 x_1^* + \cdots + \hat{\beta}_k x_k^*$. La desviación estándar estimada del estimador correspondiente es de nuevo una expresión complicada que comprende la muestra de las x_{ij}. No obstante, los mejores paquetes de estadística computarizados la calcularán cuando se les indique. Las inferencias alrededor de $\hat{\mu}_{Y \cdot x_1^*, \ldots, x_k^*}$ están basadas en estandarizar su estimador para obtener una variable t que tenga $n - (k + 1)$ grados de libertad.

1. Un intervalo de confianza $100(1 - \alpha)\%$ para β_i, el coeficiente de x_i en la función de regresión, es

$$\hat{\beta}_i \pm t_{\alpha/2, n-(k+1)} \cdot s_{\hat{\beta}_i}$$

2. Una prueba para H_0: $\beta_i = \beta_{i0}$ utiliza el valor estadístico t de $t = (\hat{\beta}_i - \beta_{i0})/s_{\hat{\beta}_i}$ basado en $n - (k + 1)$ grados de libertad. La prueba es de cola superior, cola inferior, o de dos colas, según si H_a contiene la desigualdad $>$, $<$, o $\neq$.

3. Un intervalo de confianza $100(1 - \alpha)\%$ para $\mu_{Y \cdot x_1^*, \ldots, x_k^*}$ es

$$\hat{\mu}_{Y \cdot x_1^*, \ldots, x_k^*} \pm t_{\alpha/2, n-(k+1)} \cdot \{\text{desviación estándar estimada de } \hat{\mu}_{Y x_1^*, \ldots, x_k^*}\}$$
$$= \hat{y} \pm t_{\alpha/2, n-(k+1)} \cdot s_{\hat{Y}}$$

donde $\hat{Y}$ es el estadístico $\hat{\beta}_0 + \hat{\beta}_1 x_1^* + \cdots + \hat{\beta}_k x_k^*$ y $\hat{y}$ es el valor calculado de $\hat{Y}$.

4. Un intervalo de predicción $100(1 - \alpha)\%$ para un valor futuro de y es

$$\hat{\mu}_{Y x_1^*, \ldots, x_k^*} \pm t_{\alpha/2, n-(k+1)} \cdot \{s^2 + (\text{desviación estándar estimada de } \hat{\mu}_{Y x_1^*, \ldots, x_k^*})^2\}^{1/2}$$
$$= \hat{y} \pm t_{\alpha/2, n-(k+1)} \cdot \sqrt{s^2 + s_{\hat{Y}}^2}$$

Los intervalos simultáneos, para los que la confianza simultánea o nivel de predicción es controlado, se pueden obtener al aplicar la técnica de Bonferroni.

Ejemplo 13.15 La adsorción de suelo y sedimento, la magnitud a la que se recolectan productos químicos en forma condensada en la superficie, es una importante característica que influye sobre la efectividad de plaguicidas y diversos productos químicos agrícolas. El artículo "Adsorption of Phosphate, Arsenate, Methanearsonate, and Cacodylate by Lake and Stream Sediments: Comparisons with Soils" (*J. of Environ. Qual.*, 1984: 499-504) da la información siguiente (tabla 13.5) en y = índice de adsorción de fosfato, x_1 = cantidad de hierro extraíble y x_2 = cantidad de aluminio extraíble.

Tabla 13.5 Datos para el ejemplo 13.15

Observación	x_1 = hierro extraíble	x_2 = aluminio extraíble	y = índice de adsorción
1	61	13	4
2	175	21	18
3	111	24	14
4	124	23	18
5	130	64	26
6	173	38	26
7	169	33	21
8	169	61	30
9	160	39	28
10	244	71	36
11	257	112	65
12	333	88	62
13	199	54	40

El artículo propuso el modelo

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

Un análisis de computadora dio la información siguiente:

Parámetro $\hat{\beta}_i$	Estimación $\hat{\beta}_i$	DE $s_{\hat{\beta}_i}$ estimada
β_0	−7.351	3.485
β_1	.11273	.02969
β_2	.34900	.07131

$R^2 = .948$ R^2 ajustada $= .938$ $s = 4.379$

$\hat{\mu}_{Y \cdot 160,39} = \hat{y} = -7.351 + (.11273)(160) + (.34900)(39) = 24.30$

DE estimada de $\hat{\mu}_{Y \cdot 160,39} = s_{\hat{Y}} = 1.30$

Un intervalo de confianza de 99% para β_1, el cambio en adsorción esperada asociado con un aumento de 1 unidad en hierro extraíble mientras el aluminio extraíble se mantiene fijo, requiere $t_{.005,13-(2+1)} = t_{.005,10} = 3.169$. El intervalo de confianza es

$$.11273 \pm (3.169)(.02969) = .11273 \pm .09409 \approx (.019, .207)$$

De modo semejante, un intervalo de 99% para β_2 es

$$.34900 \pm (3.169)(.07131) = .34900 \pm .22598 \approx (.123, .575)$$

La técnica de Bonferroni implica que el nivel de confianza simultáneo para ambos intervalos es al menos de 98%.

Un intervalo de confianza de 95% para $\mu_{Y \cdot 160,39}$, la adsorción esperada cuando el hierro extraíble = 160 y el aluminio extraíble = 39, es

$$24.30 \pm (2.228)(1.30) = 24.30 \pm 2.90 = (21.40, 27.20)$$

Un intervalo de predicción de 95% para un valor futuro de adsorción a observar cuando $x_1 = 160$ y $x_2 = 39$ es

$$24.30 \pm (2.228)\{(4.379)^2 + (1.30)^2\}^{1/2} = 24.30 \pm 10.18 = (14.12, 34.48)$$ ■

Es frecuente que la hipótesis de interés tenga la forma H_0: $\beta_i = 0$ para una i particular. Por ejemplo, después de ajustar el modelo de cuatro predictores del ejemplo 13.12, el investigador podría desear probar H_0: $\beta_4 = 0$. Según H_0, mientras los predictores x_1, x_2 y x_3 continúen en el modelo, x_4 no contiene información útil acerca de y. El valor del estadístico de prueba es la **razón *t*** $\hat{\beta}_i/s_{\hat{\beta}_i}$. Numerosos paquetes computarizados de estadística presentan la razón t y el correspondiente valor P para cada uno de los predictores incluidos en el modelo. Por ejemplo, la figura 13.15 muestra que mientras la potencia, temperatura y tiempo se retengan en el modelo, el predictor x_1 = fuerza se puede eliminar.

Una prueba *F* para un grupo de predictores La prueba de utilidad F del modelo era apropiada para probar si hay información útil acerca de la variable dependiente en *cualquiera* de los predictores k (es decir, si $\beta_1 = \cdots = \beta_k = 0$). En muchas situaciones, uno construye primero un modelo que contenga k predictores y luego desea saber si cualquiera de los predictores de un subconjunto particular da información útil acerca de Y. Por ejemplo, un modelo a usar para predecir calificaciones de examen de estudiantes podría incluir un grupo de variables secundarias, como son ingreso familiar y niveles de educación, y también algunas variables escolares características como el tamaño del grupo y gasto por alumno. Una hipótesis interesante es que los predictores escolares características pueden omitirse del modelo.

Los predictores se marcan como $x_1, x_2, \ldots, x_l, x_{l+1}, \ldots, x_k$ de modo que sea la última $k - l$ que se está considerando omitir del modelo. La hipótesis relevante es la siguiente:

H_0: $\beta_{l+1} = \beta_{l+2} = \cdots = \beta_k = 0$

(de modo que el modelo "reducido" $Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_l x_l + \epsilon$ es correcto) en función de

H_a: al menos una entre $\beta_{l+1}, \ldots, \beta_k$ no es 0

(de modo que en el modelo "completo" $Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon$, al menos uno de los últimos predictores $k - l$ proporciona información útil)

La prueba se realiza al ajustar tanto el modelo completo como el reducido. Debido a que el modelo completo contiene no sólo los predictores del modelo reducido sino también algunos predictores adicionales, debe ajustar los datos al menos tan bien como el modelo reducido. Esto es, si se hace que SSE_k sea la suma de residuos al cuadrado para el modelo completo y SSE_l sea la suma correspondiente para el modelo reducido, entonces $SSE_k \leq SSE_l$. Intuitivamente, si SSE_k es mucho menor que SSE_l, el modelo completo da un ajuste mucho mejor que el modelo reducido; el estadístico de prueba apropiado debe depender entonces de la reducción $SSE_l - SSE_k$ en variación inexplicada.

SSE_k = variación inexplicada para el modelo completo

SSE_l = variación inexplicada para el modelo reducido

Valor del estadístico de prueba: $$f = \frac{(SSE_l - SSE_k)/(k - l)}{SSE_k/[n - (k + 1)]} \qquad (13.20)$$

Región de rechazo: $f \geq F_{\alpha,k-l,n-(k+1)}$

Ejemplo 13.16 La información de la tabla 13.6 se tomó del artículo "Applying Stepwise Multiple Regression Analysis to the Reaction of Formaldehyde with Cotton Cellulose" (*Textile Research J.*, 1984: 157–165). La variable dependiente y es una capacidad durable de planchado, una medida cuantitativa de resistencia a las arrugas. Las cuatro variables independientes empleadas en el proceso de construcción del modelo son x_1 = concentración de HCHO (formaldehído), x_2 = razón de catalizador, x_3 = temperatura de curado, y x_4 = tiempo de curado.

Tabla 13.6 Datos para el ejemplo 13.16

Observación	x	x	x	x	y	Observación	x	x	x	x	y
1	8	4	100	1	1.4	16	4	10	160	5	4.6
2	2	4	180	7	2.2	17	4	13	100	7	4.3
3	7	4	180	1	4.6	18	10	10	120	7	4.9
4	10	7	120	5	4.9	19	5	4	100	1	1.7
5	7	4	180	5	4.6	20	8	13	140	1	4.6
6	7	7	180	1	4.7	21	10	1	180	1	2.6
7	7	13	140	1	4.6	22	2	13	140	1	3.1
8	5	4	160	7	4.5	23	6	13	180	7	4.7
9	4	7	140	3	4.8	24	7	1	120	7	2.5
10	5	1	100	7	1.4	25	5	13	140	1	4.5
11	8	10	140	3	4.7	26	8	1	160	7	2.1
12	2	4	100	3	1.6	27	4	1	180	7	1.8
13	4	10	180	3	4.5	28	6	1	160	1	1.5
14	6	7	120	7	4.7	29	4	1	100	1	1.3
15	10	13	180	3	4.8	30	7	10	100	7	4.6

Considere el modelo completo formado por $k = 14$ predictores: $x_1, x_2, x_3, x_4, x_5 = x_1^2, \ldots, x_8 = x_4^2, x_9 = x_1x_2, \ldots, x_{14} = x_3x_4$ (todos los predictores de primero y segundo órdenes). ¿Se justifica la inclusión de predictores de segundo orden? Esto es, ¿debe usarse el modelo reducido formado sólo por los predictores x_1, x_2, x_3, y x_4 ($l = 4$)? A continuación se presenta la salida que resulta de ajustar los dos modelos:

Parámetro	Estimación para modelo reducido	Estimación para modelo completo
β_0	−.9122	−8.807
β_1	.16073	.1768
β_2	.21978	.7580
β_3	.011226	.10400
β_4	.10197	.5052
β_5	—	−.04393
β_6	—	−.035887
β_7	—	−.00003271
β_8	—	−.01646
β_9	—	.00588
β_{10}	—	.002702
β_{11}	—	.01178
β_{12}	—	−.0006547
β_{13}	—	.00242
β_{14}	—	.002526
R^2	.692	.921
SSE	17.4951	4.4782

Las hipótesis a probar son

$$H_0: \beta_5 = \beta_6 = \cdots = \beta_{14} = 0$$

contra

$$H_a: \text{al menos una entre } \beta_5, \ldots, \beta_{14} \text{ no es } 0$$

Con $k = 14$ y $l = 4$, el valor crítico de F para una prueba con $\alpha = .01$ es $F_{.01,10,15} = 3.80$. El valor del estadístico de prueba es

$$f = \frac{(17.4951 - 4.4782)/10}{4.4782/15} = \frac{1.3017}{.2985} = 4.36$$

Como $4.36 \geq 3.80$, H_0 es rechazada. La conclusión es que el modelo apropiado debe incluir al menos uno de las predictores de segundo orden. ■

Evaluación de lo adecuado de un modelo

Los residuos estandarizados en regresión múltiple resultan de dividir cada uno de los residuos entre su desviación estándar estimada; la fórmula para estas desviaciones estándar es considerablemente más complicada que en el caso de regresión lineal simple. Se recomienda una gráfica de probabilidad normal de los residuos estandarizados como base para validar la suposición de normalidad. Las gráficas de residuos estandarizados en función de cada predictor y en función de $\hat{y}$ no deberían mostrar un patrón discernible. Las gráficas de residuos ajustadas también pueden ser útiles en este trabajo. El libro de Neter y otros es una referencia sumamente útil.

Ejemplo 13.17 La figura 13.16 muestra una gráfica de probabilidad normal de los residuos estandarizados para los datos de adsorción y el modelo ajustado dado en el ejemplo 13.15. La rectitud de la gráfica arroja poca duda sobre la suposición de que la desviación aleatoria está normalmente distribuida.

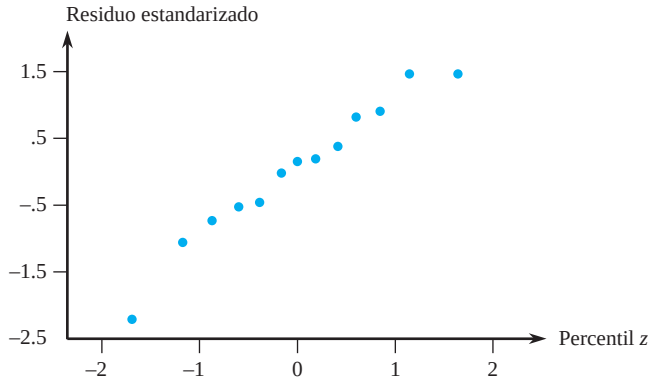


Figura 13.16 Una gráfica de probabilidad normal de los residuos estandarizados para los datos y modelo del ejemplo 13.15

La figura 13.17 muestra las otras gráficas sugeridas para los datos de adsorción. Dado que hay sólo 13 observaciones en el conjunto de datos, no hay mucha evidencia de un patrón en ninguna de las tres primeras gráficas que no sea la irregularidad. El punto situado en la parte inferior de cada una de estas tres gráficas corresponde a la observación con el residuo grande. Un poco más adelante se dirá más acerca de estas observaciones. Por ahora, no hay razón obligatoria para tomar una acción correctiva.

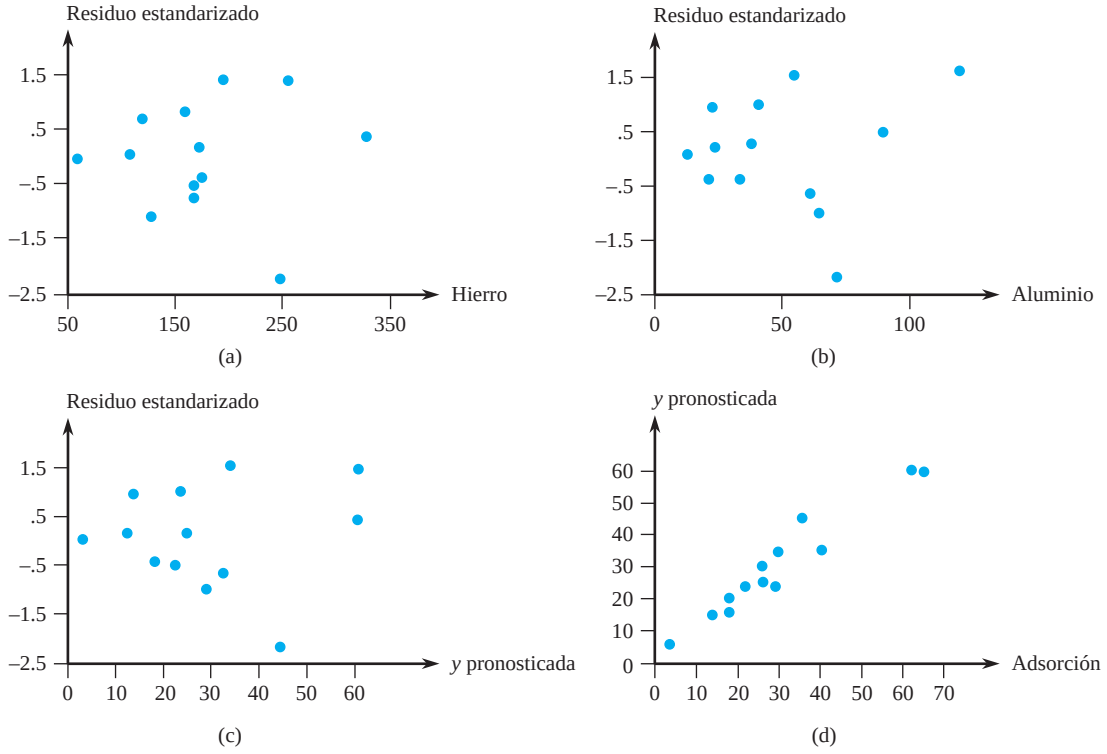


Figura 13.17 Gráficas de diagnóstico para los datos de adsorción: (a) residuos estandarizados en función de x_1; (b) residuos estandarizados en función de x_2; (c) residuos estandarizados en función de $\hat{y}$; (d) $\hat{y}$ en función de y ■

EJERCICIOS Sección 13.4 (36–54)

36. La salud cardiorrespiratoria es ampliamente reconocida como un componente importante del bienestar físico general. La medición directa de la inhalación máxima de oxígeno (VO_2 máx) es la mejor medida individual de esta salud, pero la medición directa es lenta y costosa. Por tanto, es deseable tener una ecuación de predicción para el VO_2máx en términos de cantidades que se puedan obtener con facilidad. Considere las variables

$y = VO_2$máx (L/min) $x_1 =$ peso (kg)

$x_2 =$ edad (años)

$x_3 =$ tiempo necesario para caminar 1 milla (min)

$x_4 =$ ritmo cardiaco al final de la caminata (pulsaciones/min)

He aquí un posible modelo, para estudiantes de sexo masculino, consistente con la información dada en el artículo "Validation of the Rockport Fitness Walking Test in College Males and Females" (*Research Quarterly for Exercise and Sport*, 1994: 152–158):

$$Y = 5.0 + .01x_1 - .05x_2 - .13x_3 - .01x_4 + \epsilon$$

$$\sigma = .4$$

a. Interprete β_1 y β_3.

b. ¿Cuál es el valor esperado de VO_2máx cuando el peso es de 76 kg, 20 años de edad, el tiempo de caminata es de 12 minutos y el ritmo cardiaco es de 140 p/min?

c. ¿Cuál es la probabilidad de que VO_2máx sea entre 1.00 y 2.60 para una sola observación hecha cuando los valores de los predictores sean como se expresa en el inciso (b)?

37. Una compañía de transporte por carretera consideró un modelo de regresión múltiple, para relacionar la variable dependiente y = tiempo total de viaje diario para uno de sus conductores (horas), con los predictores x_1 = distancia recorrida (millas) y x_2 = número de entregas hechas. Supóngase que la ecuación del modelo es

$$Y = -.800 + .060x_1 + .900x_2 + \epsilon$$

a. ¿Cuál es el valor medio de tiempo de viaje cuando la distancia recorrida es de 50 millas y se hacen tres entregas?

b. ¿Cómo se interpretaría $\beta_1 = .060$, el coeficiente del predictor x_1? ¿Cuál es la interpretación de $\beta_2 = .900$?

c. Si $\sigma = .5$ hora, ¿cuál es la probabilidad de que el tiempo de viaje sea a lo sumo de 6 horas cuando se hacen tres entregas y la distancia recorrida sea de 50 millas?

38. Sea y = duración de un cojinete, x_1 = viscosidad de aceite, y x_2 = carga. Supóngase que el modelo de regresión múltiple que relaciona duración con viscosidad y carga es

$$Y = 125.0 + 7.75x_1 + .0950x_2 - .0090x_1x_2 + \epsilon$$

a. ¿Cuál es el valor medio de la duración cuando la viscosidad es 40 y la carga es 1100?

b. Cuando la viscosidad es de 30, ¿cuál es el cambio en la duración media asociado con un aumento de 1 en carga? Cuando la viscosidad es de 40, ¿cuál es el cambio en la duración media asociado con un incremento de 1 en carga?

39. Sea y = ventas de un restaurante de comida rápida (miles de dólares), x_1 = número de restaurantes competidores a una milla a la redonda, x_2 = población dentro de una milla de radio (miles de personas), y x_3 es una variable indicadora igual a 1 si el restaurante tiene una ventanilla para automovilistas y 0 si no la tiene. Suponga que el modelo de regresión verdadero es

$$Y = 10.00 - 1.2x_1 + 6.8x_2 + 15.3x_3 + \epsilon$$

a. ¿Cuál es el valor medio de ventas cuando el número de restaurantes competidores es 2, hay 8000 habitantes en un radio de 1 milla, y el restaurante tiene una ventanilla para automovilistas?

b. ¿Cuál es el valor medio de ventas de un restaurante sin ventanilla para automovilistas, que tiene tres restaurantes competidores y 5000 habitantes en un radio de 1 milla?

c. Interprete β_3.

40. El artículo "Readability of Liquid Crystal Displays: A Response Surface" (*Human Factors*, 1983: 185-190) utilizó un modelo de regresión múltiple con cuatro variables independientes para estudiar la precisión en nitidez de pantallas de cristal líquido. Las variables fueron

y = porcentaje de error de cuatro dígitos para sujetos que ven una pantalla de cristal líquido

x_1 = nivel de luz de fondo (de 0 a 122 cd/m^2)

x_2 = carácter subtendido (de .025° a 1.34°)

x_3 = ángulo de visión (de 0° a 60°)

x_4 = nivel de luz ambiental (de 20 a 1500 lux)

El modelo de ajuste a los datos fue $Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \epsilon$. Los coeficientes estimados resultantes fueron $\hat{\beta}_0 = 1.52$, $\hat{\beta}_1 = .02$, $\hat{\beta}_2 = -1.40$, $\hat{\beta}_3 = .02$ y $\hat{\beta}_4 = -.0006$.

a. Calcule una estimación del porcentaje de error esperado cuando $x_1 = 10$, $x_2 = .5$, $x_3 = 50$, y $x_4 = 100$.

b. Estime el porcentaje de error medio asociado con un nivel de luz de fondo de 20, cuerda subtendida de carácter de .5, ángulo de visión de 10 y nivel de luz ambiental de 30.

c. ¿Cuál es el cambio esperado y estimado en error porcentual, cuando el nivel de luz ambiental se aumenta en 1 unidad mientras que todas las otras variables se mantienen fijas en los valores dados en el inciso (a)? Conteste para un aumento de 100 unidades en nivel de luz ambiental.

d. Explique por qué las respuestas del inciso (c) no dependen de los valores fijos de x_1, x_2 y x_3. ¿Bajo qué condiciones habría tal dependencia?

e. El modelo estimado se basó en $n = 30$ observaciones, con SST = 39.2 y SSE = 20.0. Calcule e interprete el coeficiente de determinación múltiple, y luego realice la prueba de utilidad del modelo usando $\alpha = .05$.

41. La capacidad de ecologistas para identificar regiones de máxima riqueza de especies podría tener un impacto en la preservación de la diversidad genética, que es una meta importante de la Estrategia Mundial de Conservación. El artículo "Prediction of Rarities from Habitat Variables: Coastal Plain Plants on Nova Scotian Lakeshores" (*Ecology*, 1992: 1852–1859) utilizó una muestra de $n = 37$ lagos para obtener la ecuación de regresión estimada

$$y = 3.89 + .033x_1 + .024x_2 + .023x_3 - .0080x_4 - .13x_5 - .72x_6$$

donde y = riqueza de especies, x_1 = área de cuenca de captación de aguas, x_2 = anchura de la orilla, x_3 = drenaje deficiente (%), x_4 = color del agua (total de unidades de color), x_5 = arena (%), y x_6 = alcalinidad. El coeficiente de determinación múltiple se informó como $R^2 = .83$. Realice una prueba de utilidad de modelo.

42. Una investigación de un proceso de fundición a presión produjo los datos siguientes sobre x_1 = temperatura de horno, x_2 = tiempo de cierre de matriz, y y = diferencia en temperatura en la superficie de la matriz ("A Multiple-Objective Decision-Making Approach for Assessing Simultaneous Improvement in Die Life and Casting Quality in a Die Casting Process", *Quality Engineering*, 1994: 371-383).

x_1	1250	1300	1350	1250	1300
x_2	6	7	6	7	6
y	80	95	101	85	92
x_1	1250	1300	1350	1350	
x_2	8	8	7	8	
y	87	96	106	108	

A continuación aparece la salida de Minitab por ajuste del modelo de regresión múltiple con predictores x_1 y x_2.

```
La ecuación de regresión es
diftemp = −200 + 0.210temphorno
                + 3.00 tiempocierre

Predictor         Coef        DE t-cociente      p
Constante      −199.56     11.64     −17.14 0.000
temphorno    0.210000  0.008642      24.30 0.000
tiempocierre   3.0000    0.4321       6.94 0.000

s = 1.058 R-cuadrado = 99.1% R-cuadrado (adj) = 98.8%

Análisis de varianza
FUENTE       GL       SS       MS      F      p
Regresión     2   715.50   357.75 319.31 0.000
Error         6     6.72     1.12
Total         8   722.22
```

a. Efectúe la prueba de utilidad del modelo.

b. Calcule e interprete un intervalo de confianza de 95% para β_2, el coeficiente de regresión de población de x_2.

c. Cuando $x_1 = 1300$ y $x_2 = 7$, la desviación estándar estimada de $\hat{Y}$ es $s_{\hat{Y}} = .353$. Calcule un intervalo de confianza de 95% para una verdadera diferencia promedio de temperatura, cuando la temperatura del horno sea de 1300 y el tiempo de cierre de matriz es de 7.

d. Calcule un intervalo de predicción de 95% para la diferencia de temperatura que resulte de un ciclo individual y experimental de fundición, con temperatura del horno de 1300 y tiempo de cierre de matriz de 7.

43. Un experimento realizado para estudiar el efecto del contenido molar del cobalto (x_1), y la temperatura de calcinación (x_2) en el área superficial de un catalizador de hidróxido de hierro-cobalto (y), produjo los datos siguientes ("Structural Changes and Surface Properties of $Co_xFe_{3-x}O_4$ Spinels", *J. of Chemical Tech. and Biotech.*, 1994: 161–170). Una petición al paquete SAS para ajustar $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$, donde $x_3 = x_1 x_2$ (un predictor de interacción), generó lo siguiente.

x_1	.6	.6	.6	.6	.6	1.0	1.0
x_2	200	250	400	500	600	200	250
y	90.6	82.7	58.7	43.2	25.0	127.1	112.3

x_1	1.0	1.0	1.0	2.6	2.6	2.6	2.6
x_2	400	500	600	200	250	400	500
y	19.6	17.8	9.1	53.1	52.0	43.4	42.4

x_1	2.6	2.8	2.8	2.8	2.8	2.8
x_2	600	200	250	400	500	600
y	31.6	40.9	37.9	27.5	27.3	19.0

a. Pronostique el valor del área superficial cuando el contenido de cobalto sea de 2.6 y la temperatura sea de 250, y calcule el valor del residuo correspondiente.

b. Debido a que $\hat{\beta}_1 = -46.0$, es legítimo concluir que si el contenido de cobalto aumenta en 1 unidad mientras sigan fijos los valores de los otros predictores, ¿puede esperarse que el área superficial se reduzca en alrededor de 46 unidades? Explique su razonamiento.

c. ¿Parece haber una útil relación lineal entre y y los predictores?

d. Dado que el contenido de moles y temperatura de calcinación permanecen en el modelo, ¿la interacción entre el predictor x_3 da información útil acerca de y? Exprese y pruebe las hipótesis apropiadas usando un nivel de significancia de .01.

e. La desviación estándar estimada de $\hat{Y}$, cuando el contenido de moles es de 2.0 y la temperatura de calcinación es 500, es $s_{\hat{Y}} = 4.69$. Calcule un intervalo de confianza de 95% para el valor medio del área superficial bajo estas circunstancias.

44. La salida adjunta de Minitab de regresión se basa en los datos que aparecían en el artículo "Application of Design of Experiments for Modeling Surface Roughness in Ultrasonic Vibration Turning" (*J. of Engr. Manuf.*, 2009: 641–652). La variable de respuesta es la rugosidad superficial (μm), y las variables independientes son la amplitud de vibración (mm), profundidad de corte (mm), velocidad de avance (μm/rev) y la velocidad de corte (m/min), respectivamente.

a. ¿Cuántas observaciones estaban en el conjunto de datos?

b. Interprete el coeficiente de determinación múltiple.

c. Lleve a cabo una prueba de hipótesis para decidir si el modelo especifica una relación útil entre la variable de respuesta y por lo menos uno de los predictores.

d. Interprete el número 18.2602 que aparece en la columna Coef.

e. Al nivel de significancia .10, ¿puede ser eliminada sólo una de las predictores del modelo siempre que todos los otros se conserven?

Salida SAS para el ejercicio 43

```
Variable dependiente: AREASUP

                              Análisis de varianza


 Fuente    GL Suma de cuadrados    Media cuadrática    Valor F   Prob>F
 Modelo     3      15223.52829          5074.50943     18.924    0.0001
 Error     16       4290.53971           268.15873
C Total    19      19514.06800


Raiz MSE          16.37555  R-cuadrado  0.7801
Media profundidad 48.06000    Adj R-sq  0.7389
C.V.              34.07314


                           Parámetros estimados


                      Parámetro        Error      T para H0:      Prob
Variable          GL    estimado     estándar   Parámetro = 0     > |T|
INTERCEP           1  185.485740  21.19747682           8.750    0.0001
CONTENIDO COBALTO  1  -45.969466  10.61201173          -4.332    0.0005
TEMPERATURA        1   -0.301503   0.05074421          -5.942    0.0001
CONTEMP            1    0.088801   0.02540388           3.496    0.0030
```

f. La DE estimada de $\hat{Y}$ cuando los valores de los cuatro predictores son 10, .5, .25 y 50, respectivamente, es .1178. Calcule un IC de rugosidad promedio real y un IP de la rugosidad de una sola muestra y compare los dos intervalos.

La ecuación de regresión

Ra = −0.972 − 0.0312a + 0.557d + 18.3f + 0.00282v

Predictor	Coef	DE Coef	T	P
Constante	−0.9723	0.3923	−2.48	0.015
a	−0.03117	0.01864	−1.67	0.099
d	0.5568	0.3185	1.75	0.084
f	18.2602	0.7536	24.23	0.000
v	0.002822	0.003977	0.71	0.480

S = 0.822059 R-Sq = 88.6% R-Sq(adj) = 88.0%

Fuente	GL	SS	MS	F	P
Regresión	4	401.02	100.25	148.35	0.000
Error residual	76	51.36	0.68		
Total	80	452.38			

45. El artículo "Analysis of the Modeling Methodologies for Predicting the Strength of Air-Jet Spun Yarns" (*Textile Res. J.*, 1997: 39–44) presentado en un estudio llevado a cabo para relacionar la tenacidad del hilo (y, en g/tex) con la cantidad de hilo (x_1, en tex), porcentaje de poliéster (x_2), presión de la primera tobera (x_3, en kg/cm^2), y presión de la segunda tobera (x_4, en kg/cm^2). La estimación del término constante en la correspondiente ecuación de regresión múltiple fue de 6.121. Los coeficientes estimados para los cuatro predictores fueron −.082, .113, .256 y −.219, respectivamente, y el coeficiente de determinación múltiple fue de .946.

a. Suponiendo que el tamaño muestral fue de $n = 25$, exprese y pruebe las hipótesis apropiadas para determinar si el modelo especifica una relación lineal útil entre la variable dependiente y al menos uno de los cuatro predictores del modelo.

b. Una vez más utilizando $n = 25$, calcule el valor de la R^2 ajustada.

c. Calcule un intervalo de confianza de 99% para una verdadera tenacidad media del hilo cuando la cantidad de éste es 16.5, el hilo contiene 50% de poliéster, la presión de la primera tobera es 3 y la presión de la segunda tobera es 5, si la desviación estándar estimada de tenacidad bajo estas circunstancias es de .350.

46. Un análisis de regresión efectuado para relacionar y = tiempo de reparación para un sistema de filtración de agua (h), con x_1 = tiempo transcurrido desde el servicio previo (meses) y x_2 = tipo de reparación (1 si es eléctrico y 0 si es mecánico), dio el siguiente modelo basado en $n = 12$ observaciones: $y = .950 + .400x_1 + 1.250x_2$. Además, SST = 12.72, SSE = 2.09, y $s_{\hat{\beta}_2} = .312$.

a. ¿Parece haber una relación lineal útil entre el tiempo de reparación y los dos predictores del modelo? Realice una prueba de las hipótesis apropiadas usando un nivel de significancia de .05.

b. Dado que el tiempo transcurrido desde el último servicio sigue en el modelo, ¿el tipo de reparación da información útil acerca del tiempo de reparación? Exprese y pruebe las hipótesis apropiadas usando un nivel de significancia de .01.

c. Calcule e interprete un intervalo de confianza de 95% para β_2.

d. La desviación estándar estimada de una predicción para el tiempo de reparación, cuando el tiempo transcurrido sea de 6 meses y la reparación es eléctrica, es de .192. Pronostique el tiempo de reparación bajo estas circunstancias al calcular un intervalo de predicción de 99%. ¿El intervalo sugiere que el modelo estimado dará una predicción precisa? ¿Por qué sí o por qué no?

47. El diseño eficiente de ciertos tipos de incineradores de desechos municipales exige que se disponga de información acerca del contenido energético de los desechos. Los autores del artículo "Modeling the Energy Content of Municipal Solid Waste Using Multiple Regression Analysis" (*J. of the Air and Waste Mgmnt. Assoc.*, 1996: 650-656) bondadosamente nos proporcionaron la información siguiente acerca de y = contenido energético (kcal/kg), las tres variables físicas de composición x_1 = % de plástico por peso, x_2 = % de papel por peso, y x_3 = % de basura por peso, y la variable próxima de análisis x_4 = % de humedad por peso para especímenes de desechos de cierta región.

Obs	Plástico	Papel	Basura	Agua	Contenido energético
1	18.69	15.65	45.01	58.21	947
2	19.43	23.51	39.69	46.31	1407
3	19.24	24.23	43.16	46.63	1452
4	22.64	22.20	35.76	45.85	1553
5	16.54	23.56	41.20	55.14	989
6	21.44	23.65	35.56	54.24	1162
7	19.53	24.45	40.18	47.20	1466
8	23.97	19.39	44.11	43.82	1656
9	21.45	23.84	35.41	51.01	1254
10	20.34	26.50	34.21	49.06	1336
11	17.03	23.46	32.45	53.23	1097
12	21.03	26.99	38.19	51.78	1266
13	20.49	19.87	41.35	46.69	1401
14	20.45	23.03	43.59	53.57	1223
15	18.81	22.62	42.20	52.98	1216
16	18.28	21.87	41.50	47.44	1334
17	21.41	20.47	41.20	54.68	1155
18	25.11	22.59	37.02	48.74	1453
19	21.04	26.27	38.66	53.22	1278
20	17.99	28.22	44.18	53.37	1153
21	18.73	29.39	34.77	51.06	1225
22	18.49	26.58	37.55	50.66	1237
23	22.08	24.88	37.07	50.72	1327
24	14.28	26.27	35.80	48.24	1229
25	17.74	23.61	37.36	49.92	1205
26	20.54	26.58	35.40	53.58	1221
27	18.25	13.77	51.32	51.38	1138
28	19.09	25.62	39.54	50.13	1295
29	21.25	20.63	40.72	48.67	1391
30	21.62	22.71	36.22	48.19	1372

El uso del Minitab para ajustar un modelo de regresión múltiple, con las cuatro variables citadas líneas antes como predictores de contenido energético, produjo la siguiente salida:

```
La ecuación de regresión es
contener = 2245 + 28.9 plástico
               + 7.64 papel + 4.30 basura
                −37.4 agua

Predictor        Coef     StDev         T        P
Constante      2244.9     177.9     12.62    0.000
plástico       28.925     2.824     10.24    0.000
papel           7.644     2.314      3.30    0.003
basura          4.297     1.916      2.24    0.034
agua          −37.354     1.834    −20.36    0.000

s = 31.48 R-cuadrada = 96.4% R-cuadrada (adj) = 95.8%


Análisis de varianza

Fuente       GL       SS       MS        F       P
Regresión     4   664931   166233   167.71   0.000
Error        25    24779      991
Total        29   689710
```

a. Interprete los valores de los coeficientes de regresión estimada $\hat{\beta}_1$ y $\hat{\beta}_4$.

b. Exprese y pruebe las hipótesis apropiadas para determinar si el ajuste del modelo a los datos especifica una relación lineal útil entre contenido energético y al menos uno de los cuatro predictores.

c. Dado que el % de plástico, % de papel y % de agua permanecen en el modelo, ¿el % de basura da información útil acerca del contenido energético? Exprese y pruebe las hipótesis apropiadas usando un nivel de significancia de .05.

d. Utilice el hecho de que $s_{\hat{Y}} = 7.46$ cuando $x_1 = 20$, $x_2 = 25$, $x_3 = 40$, y $x_4 = 45$ para calcular un intervalo de confianza de 95% para el verdadero contenido energético promedio bajo estas circunstancias. ¿El intervalo resultante sugiere que el contenido energético medio ha sido estimado con precisión?

e. Use la información dada en el inciso (d) para predecir el contenido energético, para una muestra de desechos que tenga las características especificadas, de modo que lleve información acerca de precisión y confiabilidad.

48. Un experimento para investigar los efectos de una nueva técnica para desengomar seda se describe en el artículo "Some Studies in Degumming of Silk with Organic Acids" (*J. Society of Dyers and Colourists*, 1992: 79–86). Una variable de respuesta de interés fue y = pérdida de peso (%). Los experimentadores hicieron observaciones de pérdida de peso para diversos valores de tres variables independientes: x_1 = temperatura (°C)= 90, 100, 110; x_2 = tiempo de tratamiento (min)= 30, 75, 120; x_3 = concentración de ácido tartárico (g/L) = 0, 8, 16. En los análisis de regresión, los tres valores de cada variable se codificaron como −1, 0 y 1, respectivamente y dieron los datos siguientes (el valor $y_8 = 19.3$ se reportó, pero el valor $y_8 = 20.3$ produjo una salida de regresión idéntica a la que aparece en el artículo).

Obs	1	2	3	4	5	6	7	8
x_1	−1	−1	1	1	−1	−1	1	1
x_2	−1	1	−1	1	0	0	0	0
x_3	0	0	0	0	−1	1	−1	1
y	18.3	22.2	23.0	23.0	3.3	19.3	19.3	20.3

Obs	9	10	11	12	13	14	15
x_1	0	0	0	0	0	0	0
x_2	−1	−1	1	1	0	0	0
x_3	−1	1	−1	1	0	0	0
y	13.1	23.0	20.9	21.5	22.0	21.3	22.6

Un modelo de regresión múltiple con $k = 9$ predictores x_1, x_2, x_3, $x_4 = x_1^2$, $x_5 = x_2^2$, $x_6 = x_3^2$, $x_7 = x_1x_2$, $x_8 = x_1x_3$ y $x_9 = x_2x_3$ se ajustó a los datos y dio por resultado $\hat{\beta}_0 = 21.967$, $\hat{\beta}_1 = 2.8125$, $\hat{\beta}_2 = 1.2750$, $\hat{\beta}_3 = 3.4375$, $\hat{\beta}_4 = -2.208$, $\hat{\beta}_5 = 1.867$, $\hat{\beta}_6 = -4.208$, $\hat{\beta}_7 = -.975$, $\hat{\beta}_8 = -3.750$, $\hat{\beta}_9 = -2.325$, SSE = 23.379, y $R^2 = .938$.

a. ¿Este modelo especifica una relación útil? Exprese y pruebe las hipótesis apropiadas usando un nivel de significancia de .01.

b. La desviación estándar estimada de $\hat{\mu}_Y$ cuando $x_1 = \cdots = x_9 = 0$ (es decir, cuando la temperatura = 100, el tiempo = 75, y la concentración = 8) es 1.248. Calcule un intervalo de confianza de 95% para la pérdida de peso esperada cuando la temperatura, el tiempo y la concentración tengan los valores especificados.

c. Calcule un intervalo de predicción de 95% para un solo valor de pérdida de peso observada cuando la temperatura, el tiempo y la concentración tengan valores de 100, 75 y 8, respectivamente.

d. El ajuste del modelo con sólo x_1, x_2 y x_3 como predictores dio $R^2 = .456$ y SSE = 203.82. ¿Al menos uno de los predictores de segundo orden proporciona información adicional de utilidad? Exprese y pruebe las hipótesis apropiadas.

49. El artículo "The Influence of Temperature and Sunshine on the Alpha-Acid Contents of Hops (*Agricultural Meteorology*, 1974: 375-382) informa de los siguientes datos en la producción (y), temperatura media del periodo entre la fecha al recibir el lúpulo y fecha de cosecharlo (x_1), y porcentaje medio de luz solar durante el mismo periodo (x_2) para una variedad de lúpulo:

x_1	16.7	17.4	18.4	16.8	18.9	17.1
x_2	30	42	47	47	43	41
y	210	110	103	103	91	76
x_1	17.3	18.2	21.3	21.2	20.7	18.5
x_2	48	44	43	50	56	60
y	73	70	68	53	45	31

A continuación se muestra una salida parcial de Minitab por el ajuste del modelo de primer orden $Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \epsilon$ empleado en el artículo:

```
Predictor        Coef         DE  t-cociente        P
Constante     415.11      82.52        5.03    0.000
Temperatura   −6.593      4.859       −1.36    0.208
Luz solar     −4.504      1.071       −4.20    0.002
s = 24.45  R-cuadrado = 76.8%  R-cuadrado(adj) = 71.6%
```

a. ¿Qué es $\hat{\mu}_{Y\cdot 18.9,43}$, y cuál es el residuo correspondiente?

b. Pruebe H_0: $\beta_1 = \beta_2 = 0$ en función de H_a: ya sea β_1 o $\beta_2 \neq 0$ al nivel de .05.

c. La desviación estándar estimada de $\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$ cuando $x_1 = 18.9$ y $x_2 = 43$ es 8.20. Use esto para obtener un intervalo de confianza de 95% para $\mu_{Y \cdot 18.9,43}$.

d. Use la información del inciso (c) a fin de obtener un intervalo de predicción de 95% para la producción en un experimento futuro cuando $x_1 = 18.9$ y $x_2 = 43$.

e. Minitab reportó que un intervalo de predicción de 95% para producción cuando $x_1 = 18$ y $x_2 = 45$ es (35.94, 151.63). ¿Cuál es un intervalo de predicción de 90% en esta situación?

f. Dado que x_2 está en el modelo, ¿retendría el lector a x_1?

g. Cuando el modelo $Y = \beta_0 + \beta_2 x_2 + \epsilon$ se ajusta, el valor resultante de R^2 es .721. Verifique que el estadístico F para probar H_0: $Y = \beta_0 + \beta_2 x_2 + \epsilon$ en función de H_a: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$ satisface $t^2 = f$, donde t es el valor del estadístico t del inciso (f).

50. **a.** Cuando el modelo $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \epsilon$ se ajusta a los datos de lúpulos del ejercicio 49, la estimación de β_5 es $\hat{\beta}_5 = .557$ con desviación estándar estimada $s_{\hat{\beta}_5} = .94$. Pruebe H_0: $\beta_5 = 0$ contra H_a: $\beta_5 \neq 0$.

b. Cada razón t $\hat{\beta}_i / s_{\hat{\beta}_i}$ $(i = 1, 2, 3, 4, 5)$ para el modelo del inciso (a) es menor a 2 en valor absoluto, pero $R^2 = .861$ para este modelo. ¿Sería correcto eliminar cada uno de los términos del modelo debido a su pequeña razón t? Explique.

c. Con el uso de $R^2 = .861$ para el modelo del inciso (a), pruebe H_0: $\beta_3 = \beta_4 = \beta_5 = 0$ (que dice que todos los términos de segundo orden se pueden eliminar).

51. El artículo "The Undrained Strength of Some Thawed Permafrost Soils" (*Canadian Geotechnical J.*, 1979: 420–427) contiene los siguientes datos sobre la resistencia al corte de suelos arenosos (y, en kPa), profundidad (x_1, en m), y contenido de agua (x_2, en %).

	y	x	x	$\hat{y}$	$y - \hat{y}$	e^*
1	14.7	8.9	31.5	23.35	−8.65	−1.50
2	48.0	36.6	27.0	46.38	1.62	.54
3	25.6	36.8	25.9	27.13	−1.53	−.53
4	10.0	6.1	39.1	10.99	−.99	−.17
5	16.0	6.9	39.2	14.10	1.90	.33
6	16.8	6.9	38.3	16.54	.26	.04
7	20.7	7.3	33.9	23.34	−2.64	−.42
8	38.8	8.4	33.8	25.43	13.37	2.17
9	16.9	6.5	27.9	15.63	1.27	.23
10	27.0	8.0	33.1	24.29	2.71	.44
11	16.0	4.5	26.3	15.36	.64	.20
12	24.9	9.9	37.8	29.61	−4.71	−.91
13	7.3	2.9	34.6	15.38	−8.08	−1.53
14	12.8	2.0	36.4	7.96	4.84	1.02

Los valores y residuos pronosticados se calcularon ajustando un modelo cuadrático completo, que produjo la función de regresión estimada

$$y = -151.36 - 16.22x_1 + 13.48x_2 + .094x_1^2 - .253x_2^2 + .492x_1x_2$$

a. Las gráficas de e^* en función de x_1, e^* en función de x_2, y e^* en función de $\hat{y}$ ¿sugieren que el modelo cuadrático completo debe modificarse? Explique su respuesta.

b. El valor de R^2 para el modelo cuadrático completo es .759. Pruebe al nivel .05 la hipótesis nula, expresando que no hay relación lineal entre la variable dependiente y cualquiera de los cinco predictores.

c. Se puede demostrar que $V(Y) = \sigma^2 = V(\hat{Y}) + V(Y - \hat{Y})$. La estimación de σ es $\hat{\sigma} = s = 6.99$ (del modelo cuadrático completo). Primero obtenga la desviación estándar estimada de $Y - \hat{Y}$ y entonces estime la desviación estándar de $\hat{Y}$ (es decir, $\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_1^2 + \hat{\beta}_4 x_2^2 + \hat{\beta}_5 x_1 x_2$) cuando $x_1 = 8.0$ y $x_2 = 33.1$. Por último, calcule un intervalo de confianza de 95% para resistencia media. [*Sugerencia:* ¿qué es $(y - \hat{y})/e^*$?]

d. El ajuste del modelo de primer orden con función de regresión $\mu_{Y \cdot x_1 x_2} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ produjo una SSE = 894.95. Pruebe al nivel .05 la hipótesis nula que expresa que todos los términos cuadráticos se pueden eliminar del modelo.

52. La utilización de sacarosa como fuente de carbono para la producción de sustancias químicas es antieconómica. La melaza de remolacha es un sustituto que se puede obtener con facilidad y es de bajo precio. El artículo "Optimization of the Production of β-Carotene from Molasses by Blakeslea Trispora (*J. of Chemical Technology and Biotechnology*, 2002: 933–943) llevó a cabo un análisis de regresión múltiple para relacionar la variable dependiente y = cantidad de β-caroteno (g/dm^3) con la cantidad de ácido linoleico de los tres predictores, cantidad de queroseno, y cantidad de antioxidante (todos en g/dm^3).

Obs	Linoleico	Queroseno	Antioxidante	Betacaro
1	30.00	30.00	10.00	0.7000
2	30.00	30.00	10.00	0.6300
3	30.00	30.00	18.41	0.0130
4	40.00	40.00	5.00	0.0490
5	30.00	30.00	10.00	0.7000
6	13.18	30.00	10.00	0.1000
7	20.00	40.00	5.00	0.0400
8	20.00	40.00	15.00	0.0065
9	40.00	20.00	5.00	0.2020
10	30.00	30.00	10.00	0.6300
11	30.00	30.00	1.59	0.0400
12	40.00	20.00	15.00	0.1320
13	40.00	40.00	15.00	0.1500
14	30.00	30.00	10.00	0.7000
15	30.00	46.82	10.00	0.3460
16	30.00	30.00	10.00	0.6300
17	30.00	13.18	10.00	0.3970
18	20.00	20.00	5.00	0.2690
19	20.00	20.00	15.00	0.0054
20	46.82	30.00	10.00	0.0640

a. El ajuste del modelo completo de segundo orden en los tres predictores dio por resultado $R^2 = .987$ y R^2 ajustada igual a .974, mientras que el ajuste del modelo de primer orden dio $R^2 = .016$. ¿Qué se concluiría acerca de los dos modelos?

b. Para $x_1 = x_2 = 30$, $x_3 = 10$, un paquete de software de estadística reportó que $\hat{y} = .66573$, $s_{\hat{Y}} = .01785$ con base en el modelo completo de segundo orden. Pronostique la cantidad de β-caroteno que resultaría de un solo ciclo experimental con los valores designados de las variables independientes, y hágalo de modo que lleve información acerca de precisión y confiabilidad.

53. Los campos nevados contienen un amplio espectro de contaminantes que pueden representar riesgos ambientales. El artículo "Atmospheric PAH Deposition: Deposition Velocities and Washout Ratios" (*J. of Environmental Engineering*, 2002: 186–195) se concentró en la precipitación de hidrocarburos poliaromáticos. Los autores propusieron un modelo de regresión múltiple para relacionar la precipitación en un tiempo especificado (y, en $\mu g/m^2$) contra dos predictores más bien complicados x_1 (μg-s/m^3) y x_2 ($\mu g/m^2$) definidos en términos de concentraciones de aire PAH para varias especies, tiempo total, y cantidad total de precipitación. A continuación aparece la información sobre fluoranteno de especies y la correspondiente salida de Minitab:

obs	x1	x2	fluoranteno
1	92017	.0026900	278.78
2	51830	.0030000	124.53
3	17236	.0000196	22.65
4	15776	.0000360	28.68
5	33462	.0004960	32.66
6	243500	.0038900	604.70
7	67793	.0011200	27.69
8	23471	.0006400	14.18
9	13948	.0004850	20.64
10	8824	.0003660	20.60
11	7699	.0002290	16.61
12	15791	.0014100	15.08
13	10239	.0004100	18.05
14	43835	.0000960	99.71
15	49793	.0000896	58.97
16	40656	.0026000	172.58
17	50774	.0009530	44.25

La ecuación de regresión es

fluoranteno = −33.5 + 0.00205 x1 + 29836 x2

Predictor	Coef	DE Coef	T	P
Constante	−33.46	14.90	−2.25	0.041
x1	0.0020548	0.0002945	6.98	0.000
x2	29836	13654	2.19	0.046

S = 44.28 R-cuadrado = 92.3% R-cuadrado(adj) = 91.2%

Análisis de varianza

Fuente	GL	SS	MS	F	P
Regresión	2	330989	165495	84.39	0.000
Error residual	14	27454	1961		
Total	16	358443			

Formule preguntas y efectúe análisis apropiados para sacar conclusiones.

54. El uso de aceros de alta resistencia (HSS) en lugar de aleaciones de aluminio y magnesio en las estructuras del cuerpo del automóvil reduce el peso del vehículo. Sin embargo, el uso de HSS sigue siendo problemático debido a las dificultades con conformabilidad limitada, el aumento de la recuperación elástica, las dificultades en formar la parte y la reducción de la vida del troquel. El artículo "Experimental Investigation of Springback Variation in Forming of High Strength Steels" (*J. of Manuf. Sci. and Engr.*, 2008: 1–9) incluyó datos sobre la recuperación elástica y = desde el ángulo de abertura de la pared y x_1 = presión del portapieza (BHP, por sus siglas en inglés). Tres diferentes proveedores de materiales y tres diferentes regímenes de lubricación (sin lubricación, el lubricante # 1 y lubricante # 2) también se utilizaron.

a. ¿Qué predictores se utilizan en un modelo para incorporar a los proveedores y la información de lubricación, además del BHP?

b. La salida adjunta de Minitab resulta de ajustar el modelo de (a) (los autores del artículo también utilizaron Minitab; graciosamente, ellos emplearon un nivel de significancia de .06 en varias pruebas de hipótesis). ¿Parece haber una relación útil entre la variable de respuesta y por lo menos uno de los indicadores? Lleve a cabo una prueba formal de hipótesis.

c. Cuando el BHP es de 1000, el material es del proveedor 1 y sin lubricación se utiliza $s_{\hat{Y}} = .524$. Calcular un 95% de IP para la elasticidad resultado de hacer una observación adicional en estas condiciones.

d. Desde la salida, parece que el régimen de lubricación no puede proporcionar información útil. Una regresión con los predictores eliminadas correspondientes da como resultado SSE = 48.426. ¿Cuál es el coeficiente de determinación múltiple para este modelo y qué se puede concluir acerca de la importancia del régimen de lubricación?

e. Un modelo con predictores de BHP, proveedor y el régimen de lubricación, así como predictores de las interacciones entre el BHP y ambos regímenes de proveedor y lubricación, se tradujo en SSE = 28.216 y $R^2 = .849$. ¿Este modelo parece mejorar en el modelo con sólo el BHP y los predictores de proveedor?

Predictor	Coef	DE Coef	T	P
Constante	21.5322	0.6782	31.75	0.000
BHP	−0.0033680	0.0003919	−8.59	0.000
Suppl_1	−1.7181	0.5977	−2.87	0.007
Suppl_2	−1.4840	0.6010	−2.47	0.019
Lub_1	−0.3036	0.5754	−0.53	0.602
Lub_2	0.8931	0.5779	1.55	0.133

S = 1.18413 R-cuadrado = 77.5% R-cuadrado(adj) = 73.8%

Fuente	GL	SS	MS	F	P
Regresión	5	144.915	28.983	20.67	0.000
Error residual	30	42.065	1.402		
Total	35	186.980			

13.5 Otros problemas en regresión múltiple

En esta sección, se tratan superficialmente varios problemas que pueden surgir cuando se efectúa un análisis de regresión múltiple. En las referencias de este capítulo consulte un tratamiento más extenso de cualquier tema particular.

Transformaciones

En ocasiones, las consideraciones teóricas sugieren una relación no lineal entre una variable dependiente y dos o más variables independientes, mientras que en otras ocasiones las gráficas de diagnóstico indican que debe usarse algún tipo de función no lineal. Es frecuente que una transformación haga lineal al modelo.

Ejemplo 13.18 Un artículo en *Lubrication Engr.*, ("Accelerated Testing of Solid Film Lubricants", 1972: 365–372) reporta sobre una investigación de la duración de lubricantes de película sólida. Se efectuaron tres conjuntos de pruebas de cojinetes en una película tipo Mil-L-8937 en cada combinación de tres cargas (3000, 6000 y 10000 psi) y tres velocidades (20, 60 y 100 rpm), y se registró la duración (horas) de cada prueba, como se muestra en la tabla 13.7.

Tabla 13.7 Datos de duración para el ejemplo 13.18

s	l(1000s)	w	s	l(1000s)	w
20	3	300.2	60	6	65.9
20	3	310.8	60	10	10.7
20	3	333.0	60	10	34.1
20	6	99.6	60	10	39.1
20	6	136.2	100	3	26.5
20	6	142.4	100	3	22.3
20	10	20.2	100	3	34.8
20	10	28.2	100	6	32.8
20	10	102.7	100	6	25.6
60	3	67.3	100	6	32.7
60	3	77.9	100	10	2.3
60	3	93.9	100	10	4.4
60	6	43.0	100	10	5.8
60	6	44.5			

El artículo contiene el comentario de que una distribución lognormal es apropiada para W, porque se sabe que $\ln(W)$ sigue una ley normal (recuerde del capítulo 4 que esto es lo que define una distribución lognormal). El modelo que aparece es $W = (c/s^a l^b) \cdot \epsilon$, del cual $\ln(W) = \ln(c) - a\ln(s) - b\ln(l) + \ln(\epsilon)$; entonces, con $Y = \ln(W)$, $x_1 = \ln(s)$, $x_2 = \ln(l)$, $\beta_0 = \ln(c)$, $\beta_1 = -a$ y $\beta_2 = -b$, se tiene un modelo de regresión lineal múltiple. Después de calcular $\ln(w_i)$, $\ln(s_i)$ y $\ln(l_i)$ para los datos, un modelo de primer orden en las variables transformadas dio los resultados que se muestran en la tabla 13.8.

Tabla 13.8 Coeficientes estimados y razones t para el ejemplo 13.18

Parámetro β_i	Estimación $\hat{\beta}_i$	DE estimada $s_{\hat{\beta}i}$	$t = \hat{\beta}_i/s_{\hat{\beta}_i}$
β_0	10.8719	.7871	13.81
β_1	−1.2054	.1710	−7.05
β_2	−1.3979	.2327	−6.01

El coeficiente de determinación múltiple (para el ajuste transformado) tiene un valor $R^2 = .781$. La función de regresión estimada para las variables transformadas es

$$\ln(w) = 10.87 - 1.21\ln(s) - 1.40\ln(l)$$

de modo que la función de regresión original se estima como

$$w = e^{10.87} \cdot s^{-1.21} \cdot l^{-1.40}$$

Se puede usar el método Bonferroni para obtener intervalos de confianza simultáneos para β_1 y β_2, y porque $\beta_1 = -a$ y $\beta_2 = -b$, los intervalos para a y b están disponibles de inmediato. ■

En la sección 13.2, el modelo de regresión logística se introdujo para relacionar una variable dicotómica y con un solo predictor. Este modelo se puede extender en una forma obvia para incorporar más de un predictor. La probabilidad de éxito p es ahora una función de los predictores $x_1, x_2, \ldots, x_k$:

$$p(x_1, \ldots, x_k) = \frac{e^{\beta_0+\beta_1 x_1+\ldots+\beta_k x_k}}{1 + e^{\beta_0+\beta_1 x_1+\ldots+\beta_k x_k}}$$

Debe utilizarse software estadístico para estimar los parámetros, calcular las desviaciones estándar pertinentes y proporcionar otra información inferencial.

Ejemplo 13.19 Los datos fueron obtenidos de 189 mujeres que dieron a luz durante un periodo determinado en el Centro Médico de Bayside en Springfield, MA, con el fin de identificar los factores asociados con el bajo peso al nacer. La salida adjunta de Minitab es el resultado de una regresión logística en la que la variable dependiente indica si (1) o no (0) un niño con peso bajo al nacer (<2500 g) y los predictores fueron el peso de la madre en su último periodo menstrual, la edad de la madre y una variable indicadora de si (1) o no (0) la madre había fumado durante el embarazo.

Tabla de regresión logística

Predictor	Coef	DE Coef	Z	P	Cociente de posibilidades	95% Inferior	IC Superior
Constante	2.06239	1.09516	1.88	0.060			
Peso	−0.01701	0.00686	−2.48	0.013	0.98	0.97	1.00
Edad	−0.04478	0.03391	−1.32	0.187	0.96	0.89	1.02
Fumadora	0.65480	0.33297	1.97	0.049	1.92	1.00	3.70

Parece que la edad no es un importante predictor de bajo peso al nacer, a condición de que los otros dos predictores se conserven. Los otros dos parecen ser de carácter informativo. La estimación puntual de la razón impar asociada con el consumo de tabaco es 1.92 [cociente de las probabilidades de bajo peso al nacer para una fumadora y las probabilidades de una persona no fumadora, donde las *posibilidades* $= P(Y = 1)/P(Y = 0)$], en el 95% nivel de confianza, las posibilidades de un niño de bajo peso al nacer podrían ser hasta 3.7 veces más alta para una fumadora que para una persona que no fuma. ■

Por favor, consulte una de las referencias del capítulo para obtener más información sobre la regresión logística, incluidos los métodos para evaluar la eficacia y la adecuación del modelo.

Variables estandarizadoras

En la sección 13.3 se consideró transformar x en $x' = x - \bar{x}$ antes de ajustar un polinomio. Para regresión múltiple, en especial cuando los valores de variables son grandes en magnitud, es ventajoso adelantar un paso más esta codificación. Sean $\bar{x}_i$ y s_i el promedio muestral y la desviación estándar muestral de las x_{ij} $(j = 1, \ldots, n)$. Ahora se codifica cada variable x_i por $x_i' = (x_i - \bar{x}_i)/s_i$. La variable codificada x_i' simplemente vuelve a expresar cualquier valor x_i en unidades de desviación estándar arriba o debajo de la media. Entonces, si $\bar{x}_i = 100$ y $s_i = 20$, $x_i = 130$ se convierte en $x_i' = 1.5$ porque 130 está 1.5 desviaciones estándar arriba de la media de los valores de x_i. Por ejemplo, el modelo completo codificado de segundo orden con dos variables independientes tiene función de regresión.

$$E(Y) = \beta_0 + \beta_1\left(\frac{x_1 - \bar{x}_1}{s_1}\right) + \beta_2\left(\frac{x_2 - \bar{x}_2}{s_2}\right) + \beta_2\left(\frac{x_1 - \bar{x}_1}{s_1}\right)^2$$
$$+ \beta_4\left(\frac{x_2 - \bar{x}_2}{s_2}\right)^2 + \beta_5\left(\frac{x_1 - \bar{x}_1}{s_1}\right)\left(\frac{x_2 - \bar{x}_2}{s_2}\right)$$
$$= \beta_0 + \beta_1 x_1' + \beta_2 x_2' + \beta_3 x_3' + \beta_4 x_4' + \beta_5 x_5'$$

Los beneficios de codificar son (1) precisión numérica mejorada en todos los cálculos, y (2) estimación más precisa que para los parámetros del modelo no codificado, porque los parámetros individuales del modelo codificado caracterizan el comportamiento de la función de regresión cerca del centro de los datos, en lugar de cerca del origen.

Ejemplo 13.20 El artículo "The Value and the Limitations of High-Speed Turbo-Exhausters for the Removal of Tar-Fog from Carburetted Water-Gas" (*J. of Soc. Chemical Industry*, 1946: 166–168) presenta los datos (en la tabla 13.9) sobre y = contenido de alquitrán (granos/100 pie^3) de una corriente de gas como función de x_1 = velocidad del rotor (rpm) y x_2 = temperatura del gas de entrada (°F). La información también está considerada en el artículo "Some Aspects of Nonorthogonal Data Analysis" (*J. of Quality Technology*, 1973: 67–79), que sugiere el uso del modelo codificado descrito previamente.

Tabla 13.9 Datos para el ejemplo 13.20

Prueba	y	x	x	x_1'	x_2'
1	60.0	2400	54.5	−1.52428	−.57145
2	61.0	2450	56.0	−1.39535	−.35543
3	65.0	2450	58.5	−1.39535	.00461
4	30.5	2500	43.0	−1.26642	−2.22763
5	63.5	2500	58.0	−1.26642	−.06740
6	65.0	2500	59.0	−1.26642	.07662
7	44.0	2700	52.5	−.75070	−.85948
8	52.0	2700	65.5	−.75070	1.01272
9	54.5	2700	68.0	−.75070	1.37276
10	30.0	2750	45.0	−.62177	−1.93960
11	26.0	2775	45.5	−.55731	−1.86759
12	23.0	2800	48.0	−.49284	−1.50755
13	54.0	2800	63.0	−.49284	.65268
14	36.0	2900	58.5	−.23499	.00461
15	53.5	2900	64.5	−.23499	.86870
16	57.0	3000	66.0	.02287	1.08472
17	33.5	3075	57.0	.21627	−.21141
18	34.0	3100	57.5	.28073	−.13941
19	44.0	3150	64.0	.40966	.79669
20	33.0	3200	57.0	.53859	−.21141
21	39.0	3200	64.0	.53859	.79669
22	53.0	3200	69.0	.53859	1.51677
23	38.5	3225	68.0	.60305	1.37276
24	39.5	3250	62.0	.66752	.50866
25	36.0	3250	64.5	.66752	.86870
26	8.5	3250	48.0	.66752	−1.50755
27	30.0	3500	60.0	1.31216	.22063
28	29.0	3500	59.0	1.31216	.07662
29	26.5	3500	58.0	1.31216	−.06740
30	24.5	3600	58.0	1.57002	−.06740
31	26.5	3900	61.0	2.34360	.36465

Las medias y desviaciones estándar son $\bar{x}_1 = 2991.13, s_1 = 387.81, \bar{x}_2 = 58.468$, y $s_2 = 6.944$, así $x_1' = (x_1 - 2991.13)/387.81$ y $x_2' = (x_2 - 58.468)/6.944$. Con $x_3' = (x'_1)^2, x_4' = (x_2')^2, x_5' = x_1' \cdot x_2'$, el ajuste del modelo completo de segundo orden dio $\hat{\beta}_0 = 40.2660, \hat{\beta}_1 = -13.4041, \hat{\beta}_2 = 10.2553, \hat{\beta}_3 = 2.3313, \hat{\beta}_4 = -2.3405$ y $\hat{\beta}_5 = 2.5978$. La ecuación de regresión estimada es entonces

$$\hat{y} = 40.27 - 13.40x_1' + 10.26x_2' + 2.33x_3' - 2.34x_4' + 2.60x_5'$$

Entonces, si $x_1 = 3200$ y $x_2 = 57.0, x_1' = .539, x_2' = -.211, x_3' = (.539)^2 = .2901$, $x_4' = (-.211)^2 = .0447$, y $x_5' = (.539)(-.211) = -.1139$, de modo que

$$\hat{y} = 40.27 - (13.40)(.539) + (10.26)(-.211) + (2.33)(.2901) - (2.34)(.0447) + (2.60)(-.1139) = 31.16$$ ■

Selección de variable

Supongamos que un investigador ha obtenido los datos en una variable y de respuesta, así como sobre los p candidatos predictores $x_1, \ldots, x_p$. ¿Cómo puede un mejor modelo (en algún sentido) implicar un subconjunto de estos predictores para que sea seleccionado? Recordemos que los predictores se añaden uno por uno en un modelo, la SSE no puede aumentar (un modelo más grande no puede explicar una menor variación que uno más pequeño) y se reducirá por lo general, aunque tal vez por una pequeña cantidad. Así que no hay misterio en cuanto a qué modelo le da el mayor valor de R^2, debe ser el que contiene todos los predictores p. Lo que realmente nos gustaría es un modelo de participación con relativamente pocos predictores que sea fácil de interpretar y utilizar, sin embargo, explicando una cantidad relativamente grande de la variación y observada.

Para cualquier número fijo de predictores (por ejemplo, 5), es razonable para identificar el mejor modelo de ese tamaño como el que tiene el mayor valor de R^2, equivalentemente, el menor valor de SSE. La cuestión más difícil se refiere a la selección de un criterio que permita la comparación de modelos de diferentes tamaños. Vamos a usar un subíndice k para denotar una cantidad calculada a partir de un modelo que contiene k predictores (por ejemplo, SSE_k). Tres criterios diferentes, cada uno una función simple de SSE_k, son ampliamente utilizados.

1. R_k^2, el coeficiente de determinación múltiple de un modelo k-predictor. Debido a que R_k^2 casi siempre aumenta a medida que k lo hace (y nunca puede disminuir), no estamos interesados en el k que maximiza R_k^2. En su lugar, queremos identificar una pequeña k para el que R_k^2 es casi tan grande como R^2 para todos los predictores del modelo.
2. $\text{MSE}_k = \text{SSE}_k/(n - k - 1)$, el error medio cuadrático de un modelo k-predictor. Esto es a menudo usado en lugar de R_k^2, porque aunque R_k^2 no disminuye con el aumento de k, una pequeña disminución en SSE_k obtenida con un predictor adicional puede ser más que compensado por una disminución de 1 en el denominador de MSE_k. El objetivo es entonces encontrar el modelo que tiene el mínimo MSE_k. Dado que $R_k^2 = 1 - \text{MSE}_k/\text{MST}$ es *ajustado*, donde $\text{MST} = \text{SST}/(n-1)$ es constante en k, la revisión de R_k^2 *ajustado* equivale a la consideración de MSE_k.
3. La razón para el tercer criterio, C_k, es más difícil de entender, pero el criterio es ampliamente utilizado por los analistas de datos. Supongamos que el modelo de regresión real es especificado por m predictores, es decir,

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m + \epsilon \quad V(\epsilon) = \sigma^2$$

de manera que

$$E(Y) = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m$$

Considere ajustar un modelo mediante el uso de un subconjunto de k de estos m predictores; para más sencillez, suponga que se usa $x_1, x_2, \ldots, x_k$. Entonces al resolver el sistema de ecuaciones normales, se obtienen estimaciones $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ (pero no, por supuesto, estimaciones de ninguna de las β correspondientes a predictores que no estén en el modelo ajustado). El verdadero valor esperado $E(Y)$ puede entonces ser estimado por $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k$. Ahora considere el **error total de estimación esperado normalizado**

$$\Gamma_k = \frac{E\left(\sum_{i=1}^{n} [\hat{Y}_i - E(Y_i)]^2\right)}{\sigma^2} = \frac{E(\text{SSE}_k)}{\sigma^2} + 2(k+1) - n \tag{13.21}$$

La segunda igualdad en (13.21) debe tomarse de buena fe porque requiere un argumento complicado de valor esperado. Un subconjunto particular es entonces atractivo si su valor Γ_k es pequeño. Desafortunadamente, sin embargo, $E(\text{SSE}_k)$ y σ^2 no se conocen. Para solucionar esto, se denota con s^2 la estimación de σ^2 con base en el modelo que incluye todos los predictores para los que se dispone de información y se define

$$C_k = \frac{\text{SSE}_k}{s^2} + 2(k+1) - n$$

Un modelo deseable es entonces especificado por un subconjunto de predictores para los que C_k es pequeña.

El número total de modelos que se pueden crear a partir de los predictores en el grupo de candidatos es 2^p (porque cada predictor se puede incluir dentro o al margen de cualquier modelo, uno de ellos es el modelo que no contiene predictores). Si $p \leq 5$, entonces no sería demasiado tedioso examinar todos los posibles modelos de regresión que implican estos predictores usando cualquier buen paquete de software estadístico. Pero el esfuerzo computacional requerido para todos los modelos posibles se vuelve prohibitivo como el tamaño de los incrementos de la lista de candidatos. Varios paquetes de software han incorporado algoritmos que tamizan a través de modelos de diferentes tamaños con el fin de identificar mejor uno o más modelos de cada tamaño en particular. Minitab, por ejemplo, lo hará para $p \leq 31$ y permite al usuario especificar el número de modelos de cada tamaño (1, 2, 3, 4 o 5) que serán identificados con mejores valores de criterio. Usted podría preguntarse por qué nos gustaría ir más allá del mejor modelo único de cada tamaño. La respuesta es que el 2° o 3° mejor modelo puede ser más fácil de interpretar y utilizar lo que sería el mejor modelo, o puede ser más satisfactorio desde el punto de vista de adecuación del modelo. Por ejemplo, supongamos que la lista de candidatos incluye todos los predictores de un modelo cuadrático completo a partir de cinco variables independientes. Entonces, el mejor modelo predictor en 3 podría tener x_2, x_4^2 y $x_3 x_5$ predictores, mientras que el segundo mejor modelo podría ser el único con predictores x_2, x_3 y $x_2 x_3$.

Ejemplo 13.21 El artículo de repaso de Ron Hocking citado en la bibliografía del capítulo reporta un análisis de datos tomados de las ediciones de 1974 de la revista *Motor Trend*. La variable dependiente y fue el rendimiento de combustible, hubo $n = 32$ observaciones y los predictores para los que se obtuvo información fueron x_1 = forma del motor (1 = en línea y 0 = en V), x_2 = número de cilindros, x_3 = tipo de transmisión (1 = manual y 0 = auto), x_4 = número de velocidades de la transmisión, x_5 = tamaño del motor, x_6 = potencia, x_7 = número de gargantas del carburador, x_8 = relación final de impulsión, x_9 = peso y x_{10} = tiempo en ¼ de milla. En la tabla 13.10, se presenta información breve del análisis. La tabla describe, para cada k, el subconjunto que tenga una SSE_k mínima; hacia abajo en la columna de las variables se indica cuál variable se agrega al pasar de k a $k + 1$ (al pasar de $k = 2$ a $k = 3$ se suman x_3 y x_{10}, y x_2 se elimina). La figura 13.18 contiene gráficas de R_k^2, R_k^2 ajustada, y C_k en función de k; estas gráficas son una ayuda visual importante al seleccionar un subconjunto. La estimación de σ^2 es $s^2 = 6.24$, que es MSE_{10}. Un modelo

sencillo que se clasifica alto según todos los criterios es el que contiene los predictores x_3, x_9 y x_{10}.

Tabla 13.10 Mejores subconjuntos para datos de rendimiento de combustible del ejemplo 13.21

k = Número de predictores	Variables	SSE_k	R_k^2	R_k^2 ajustada	C_k
1	9	247.2	.756	.748	11.6
2	2	169.7	.833	.821	1.2
3	3, 10, −2	150.4	.852	.836	.1
4	6	142.3	.860	.839	.8
5	5	136.2	.866	.840	1.8
6	8	133.3	.869	.837	3.4
7	4	132.0	.870	.832	5.2
8	7	131.3	.871	.826	7.1
9	1	131.1	.871	.818	9.0
10	2	131.0	.871	.809	11.0

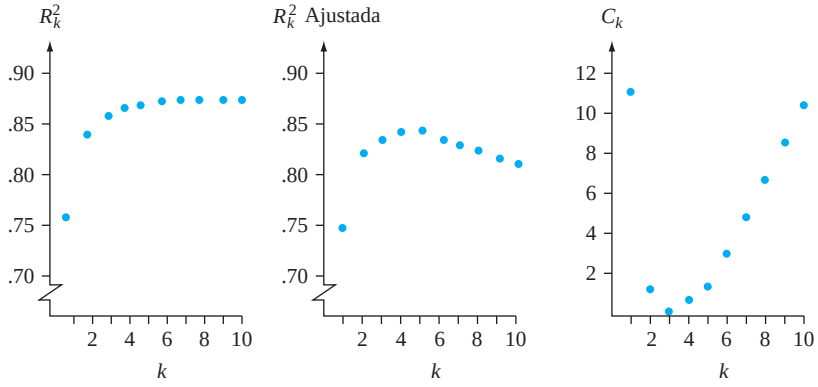


Figura 13.18 Gráficas de R_k^2 y C_k para los datos de rendimiento de combustible ■

En términos generales, cuando un subconjunto de k predictores ($k < m$) se usa para ajustar un modelo, los estimadores $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ estarán sesgados por $\beta_0, \beta_1, \ldots, \beta_k$ y $\hat{Y}$ también será un estimador sesgado para la verdadera $E(Y)$ (todo esto porque $m - k$ predictores faltan en el modelo ajustado). No obstante, según sean medidas por el error esperado normalizado total Γ_k, las estimaciones basadas en un subconjunto pueden dar más precisión de la que se obtendría si se usan todos los predictores posibles; en esencia, esta mayor precisión se obtiene al precio de introducir un sesgo en los estimadores. Un valor de k para el que $C_k \approx k + 1$ indica que el sesgo asociado con este modelo de k predictores será pequeño.

Ejemplo 13.22 La información de la resistencia de pegamento al corte, introducida en el ejemplo 13.12, contiene valores de cuatro variables independientes diferentes $x_1 - x_4$. Se encuentra que el modelo con sólo estas cuatro variables como predictores fue útil, y no hay razón obligatoria para considerar la inclusión de predictores de segundo orden. La figura 13.19 es la salida de Minitab que resulta de una petición para identificar los dos mejores modelos de cada tamaño dado.

El mejor modelo de dos predictores, con predictores de potencia y temperatura, parece ser una muy buena opción en todas las cantidades: R^2 es considerablemente más alta que para modelos con menos predictores pero casi tan grande como en modelos más grandes, R^2 ajustada está casi a su máximo para estos datos, y C_2 es pequeña y cercana a $2 + 1 = 3$.

Respuesta es resistencia

Variables	R-cuadrado	Adj. R-cuadrado	C-p	s	fuerza	potencia	temperatura	tiempo
1	57.7	56.2	11.0	5.9289		X		
1	10.8	7.7	51.9	8.6045			X	
2	68.5	66.2	3.5	5.2070		X	X	
2	59.4	56.4	11.5	5.9136		X		X
3	70.2	66.8	4.0	5.1590		X	X	X
3	69.7	66.2	4.5	5.2078	X	X	X	
4	71.4	66.8	5.0	5.1580	X	X	X	X

Figura 13.19 Salida de la opción Mejores Subconjuntos de Minitab ■

Regresión por pasos Cuando el número de predictores es demasiado grande para tener en cuenta el examen explícito o implícito de todos los subconjuntos posibles, varios procedimientos de selección alternativos por lo general identificarán buenos modelos. El procedimiento más sencillo es el método de **eliminación inversa** (BE). Este método empieza con el modelo en el que se usan todos los predictores bajo consideración. Sea el conjunto de todos los predictores $x_1, \ldots, x_m$. Entonces se examina cada una de las razones $t\ \hat{\beta}_1/s_{\hat{\beta}_i} (i = 1, \ldots, m)$ apropiada para probar H_0: $\beta_i = 0$ en función de H_a: $\beta_i \neq 0$. Si la razón t con el valor absoluto más pequeño es menor que una constante especificada previamente t_{sal}, es decir, si

$$\min_{i=1,\ldots,m} \left| \frac{\hat{\beta}_i}{s_{\hat{\beta}_i}} \right| < t_{\text{sal}}$$

entonces el predictor que corresponde a la razón más pequeña se elimina del modelo. El modelo reducido se ajusta ahora, se examinan de nuevo las $m - 1$ razones t y se elimina otro predictor si corresponde a la razón t absoluta más pequeña que t_{sal}. En esta forma, el algoritmo continúa hasta que, en alguna etapa, todas las razones t absolutas son al menos t_{sal}. El modelo utilizado es el que contiene todas los predictores que no fueron eliminados. Es frecuente que el valor $t_{\text{sal}} = 2$ se recomiende porque casi todos los valores $t_{,05}$ son cercanos a 2. Algunos paquetes de software se concentran en valores P más que en razones t.

Ejemplo 13.23 (Continuación del ejemplo 13.20)

Para el modelo cuadrático completo codificado en el que y = contenido de alquitrán, los cinco predictores potenciales son x_1', x_2', $x_3' = x_1'^2$, $x_4' = x_2'^2$ y $x_5' = x_1'x_2'$ (de modo que $m = 5$). Sin especificar t_{sal}, el predictor con la razón t absoluta más pequeña (con asterisco) se eliminó en cada etapa, produciendo la secuencia de modelos que se muestra en la tabla 13.11.

Tabla 13.11 Resultados de eliminación inversa para los datos del ejemplo 13.20

		|razón t|				
Paso	**Predictores**	**1**	**2**	**3**	**4**	**5**
1	1, 2, 3, 4, 5	16.0	10.8	2.9	2.8	1.8*
2	1, 2, 3, 4	15.4	10.2	3.7	2.0*	—
3	1, 2, 3	14.5	12.2	4.3*	—	—
4	1, 2	10.9	9.1*	—	—	—
5	1	4.4*	—	—	—	—

Con el uso de $t_{\text{sal}} = 2$, el modelo resultante estaría basado en x_1', x_2' y x_3', puesto que en el paso 3 no podría eliminarse ningún predictor. Puede verificarse que cada subconjunto es

en realidad el mejor subconjunto de su tamaño, aunque bajo ninguna circunstancia éste siempre sea el caso. ■

Una alternativa al procedimiento de eliminación inversa es la **selección directa** (FS). FS empieza sin predictores en el modelo y considera ajustar a su vez el modelo con sólo x_1, sólo x_2, . . . , y finalmente sólo x_m. La variable que, cuando se ajusta, da la razón t absoluta más grande, entra al modelo siempre que la razón rebase a la constante especificada t_{ent}. Suponga que x_1 entra al modelo. Entonces los modelos con (x_1, x_2), (x_1, x_3), . . . (x_1, x_m) se consideran a su vez. La $|\hat{\beta}_j/s_{\hat{\beta}_j}|(j = 2, \ldots, m)$ más grande especifica entonces el predictor entrante siempre que este máximo también exceda a t_{ent}. Esto continúa hasta que en algún paso ninguna de las razones t absolutas exceden de t_{ent}. Los predictores introducidos especifican entonces el modelo. El valor $t_{ent} = 2$ se usa con frecuencia por la misma razón que $t_{sal} = 2$ se usa en eliminación inversa (BE). Para los datos de contenido de alquitrán, la selección directa (FS) produjo la secuencia de modelos dada en los pasos 5, 4, . . . , 1 en la tabla 13.11 y por tanto está de acuerdo con la BE. Éste no siempre será el caso.

El procedimiento por pasos de más uso es una combinación de FS y BE, denotado por FB. Este procedimiento empieza igual que la selección directa, agregando variables al modelo, pero después de cada adición examina las variables previamente introducidas para ver si cualquiera de ellas es candidata a eliminarse. Por ejemplo, si hay ocho predictores bajo consideración y el conjunto actual consta de x_2, x_3, x_5 y x_6 con x_5 acabando de ser agregada, se examinan las razones t, $\hat{\beta}_2/s_{\hat{\beta}_2}$, $\hat{\beta}_3/s_{\hat{\beta}_3}$ y $\hat{\beta}_6/s_{\hat{\beta}_6}$. Si la razón absoluta más pequeña es menor que t_{sal}, entonces se elimina del modelo la variable correspondiente (algunos paquetes de software basan su decisión en $f = t^2$). La idea que hay detrás de FB es que, con selección directa, una sola variable puede estar más fuertemente relacionada con y que con cualquiera de las dos o más variables individualmente, pero la combinación de estas variables puede hacer que con posterioridad la variable individual sea redundante. Esto ocurrió en realidad con los datos de rendimiento de combustible que se vio en el ejemplo 13.21, con x_2 entrando y subsecuentemente saliendo del modelo.

Aun cuando en casi todas las situaciones estos procedimientos de selección automática identificarán un buen modelo, no hay garantía de que resulte el mejor modelo o incluso uno que se le aproxime a éste. Debe hacerse un escrutinio minucioso de los conjuntos de datos para los que parece haber fuertes relaciones entre algunos de los potenciales predictores; en breve se tratará más de esto.

Identificación de observaciones influyentes

En regresión lineal simple, es fácil ubicar una observación cuyo valor x sea mucho mayor o mucho menor que otros valores x de la muestra. Esta observación puede tener un gran impacto en la ecuación de regresión estimada (si en realidad depende de qué tan lejos se encuentre el punto (x, y) de la recta determinada por los otros puntos en la gráfica de dispersión**)**. En regresión múltiple, también es deseable saber si los valores de los predictores para una observación particular son tales que tiene el potencial para ejercer gran influencia en la ecuación estimada. Un método para identificar observaciones potencialmente influyentes se apoya en el hecho de que como cada $\hat{\beta}_i$ es una función lineal de $y_1, y_2, \ldots, y_n$, cada valor y pronosticado de la forma $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k$ también es una función lineal de las y_j. En particular, los valores pronosticados correspondientes a observaciones muestrales se pueden escribir como sigue:

$$\begin{aligned}
\hat{y}_1 &= h_{11}y_1 + h_{12}y_2 + \cdots + h_{1n}y_n \\
\hat{y}_2 &= h_{21}y_1 + h_{22}y_2 + \cdots + h_{2n}y_n \\
&\vdots \quad\quad \vdots \quad\quad\quad \vdots \quad\quad\quad\quad \vdots \\
\hat{y}_n &= h_{n1}y_1 + h_{n2}y_2 + \ldots + h_{nn}y_n
\end{aligned}$$

Cada uno de los coeficientes h_{ij} es una función de las x_{ij} de la muestra y no de las y_j. Se puede demostrar que $h_{ij} = h_{ji}$ y que $0 \leq h_{jj} \leq 1$.

Hay que concentrarse en los coeficientes "diagonales" $h_{11}, h_{22}, \ldots, h_{nn}$. El coeficiente h_{jj} es el peso dado a y_j al calcular el correspondiente valor $\hat{y}_j$ pronosticado. Esta cantidad también se puede expresar como una medida de la distancia entre el punto $(x_{1j}, \ldots, x_{jk})$ en espacio k-dimensional y el centro de los datos $(\bar{x}_1, \ldots, \bar{x}_k)$. Por tanto, es natural caracterizar una observación cuya h_{jj} es relativamente grande como una que tiene influencia potencialmente grande. A menos que haya una relación lineal perfecta entre los k predictores, $\sum_{j=1}^{n} h_{jj} = k + 1$, así el promedio de las h_{jj} es $(k + 1)/n$. Algunos estadísticos sugieren que si $h_{jj} > 2(k + 1)/n$, la j-ésima observación se cite como potencialmente influyente; otros usan $3(k + 1)/n$ como la línea divisoria.

Ejemplo 13.24 Los datos siguientes aparecieron en el artículo "Testing for the Inclusion of Variables in Linear Regression by a Randomization Technique" (*Technometrics,* 1966: 695–699) y fue reanalizada en la obra de Hoaglin y Welsch, "The Hat Matrix in Regression and ANOVA" (*Amer. Statistician*, 1978: 17–23). Las h_{ij} (con elementos debajo de la diagonal omitidos por simetría) siguen a los datos.

Número de vigueta	Gravedad específica (x)	Contenido de humedad (x)	Resistencia (x)
1	.499	11.1	11.14
2	.558	8.9	12.74
3	.604	8.8	13.13
4	.441	8.9	11.51
5	.550	8.8	12.38
6	.528	9.9	12.60
7	.418	10.7	11.13
8	.480	10.5	11.70
9	.406	10.5	11.02
10	.467	10.7	11.41

	1	2	3	4	5	6	7	8	9	10
1	.418	−.002	.079	−.274	−.046	.181	.128	.222	.050	.242
2		.242	.292	.136	.243	.128	−.041	.033	−.035	.004
3			.417	−.019	.273	.187	−.126	.044	−.153	.004
4				.604	.197	−.038	.168	−.022	.275	−.028
5					.252	.111	−.030	.019	−.010	−.010
6						.148	.042	.117	.012	.111
7							.262	.145	.277	.174
8								.154	.120	.168
9									.315	.148
10										.187

Aquí, $k = 2$ así que $(k + 1)/n = 3/10 = .3$; como $h_{44} = .604 > 2(.3)$, el cuarto punto de datos se identifica como potencialmente influyente. ■

Otra técnica para evaluar la influencia de la j-ésima observación, que toma en cuenta y_j así como los valores predictores, comprende eliminar la j-ésima observación del conjunto de datos y efectuar una regresión con base en las observaciones restantes. Si los coeficientes estimados de la regresión de "observación borrada" difieren en gran medida de las estimaciones basadas en los datos completos, la j-ésima observación ha tenido claramente un impacto considerable en el ajuste. Una forma de juzgar si los coeficientes estimados cambian grandemente es expresar cada cambio con relación a la desviación estándar estimada del coeficiente:

$$\frac{(\hat{\beta}_i \text{ antes de la eliminacion}) - (\hat{\beta}_i \text{ después de la eliminacion})}{s_{\hat{\beta}_i}} = \frac{\text{cambio en } \hat{\beta}_i}{s_{\hat{\beta}_i}}$$

Existen fórmulas computacionales eficientes que permiten obtener toda esta información de la regresión "sin eliminar", de modo que otras n regresiones no son necesarias.

Ejemplo 13.25 (Continuación del ejemplo 13.24)

Considere separadamente borrar las observaciones 1 y 6, cuyos residuos son los más grandes, y la observación 4, donde h_{jj} es grande. La tabla 13.12 contiene la información relevante.

Tabla 13.12 Cambios en coeficientes estimados para el ejemplo 13.25

			Cambio cuando el punto j se borra		
Parámetro	**Estimaciones sin borrar**	**DE estimada**	$j = 1$	$j = 4$	$j = 6$
β_0	10.302	1.896	2.710	−2.109	−.642
β_1	8.495	1.784	−1.772	1.695	.748
β_2	.2663	.1273	−.1932	.1242	.0329
		e_j:	−3.25	−.96	2.20
		h_{jj}:	.418	.604	.148

Para borrar el punto 1 y el punto 4, el cambio en cada estimación está en el rango de 1–1.5 desviaciones estándar, que es razonablemente importante (esto no nos dice qué ocurriría si ambos puntos se omitieran al mismo tiempo.) Para el punto 6, no obstante, el cambio es casi .25 de una desviación estándar. Por tanto, los puntos 1 y 4, pero no el 6, bien podrían omitirse al calcular una ecuación de regresión. ■

Multicolinealidad

En numerosos conjuntos de datos de regresión múltiple, $x_1, x_2, \ldots, x_k$ son altamente interdependientes. Considere el modelo usual

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon$$

con datos $(x_{1j}, \ldots, x_{kj}, y_j)$ $(j = 1, \ldots, n)$ disponibles para ajuste. Si se usa el principio de mínimos cuadrados para hacer regresión de x_i en los otros predictores $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k$, se obtiene

$$\hat{x}_i = a_0 + a_1 x_1 + \cdots + a_{i-1} x_{i-1} + a_{i+1} x_{i+1} + \cdots + a_k x_k$$

Se puede demostrar que

$$V(\hat{\beta}_i) = \frac{\sigma^2}{\sum_{j=1}^{n} (x_{ij} - \hat{x}_{ij})^2} \tag{13.22}$$

Cuando los valores muestrales x_i se pueden predecir muy bien a partir de otros valores de pronóstico, el denominador de (13.22) será pequeño, de modo que $V(\hat{\beta}_i)$ será muy grande. Si éste es el caso para al menos un predictor, se dice que la información exhibe **multicolinealidad.** Es frecuente que la multicolinealidad sea sugerida por una salida computarizada de regresión en la que R^2 es grande, pero algunas de las razones t, $\hat{\beta}_i / s_{\hat{\beta}_i}$ son pequeñas para predictores que, basados en información previa e intuición, parecen importantes. Otro indicio de la presencia de multicolinealidad está en un valor $\hat{\beta}_i$ que tiene el signo contrario de aquel que sugeriría la intuición, lo que indica que otro predictor o conjunto de predictores está sirviendo como "apoderado" de x_i.

Puede obtenerse una evaluación de la magnitud de la multicolinealidad si se hace regresión de cada predictor a la vez en los $k - 1$ predictores restantes. Denote con R_i^2 el valor de la R^2 en la regresión con variable dependiente x_i y predictores $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k$. Se ha sugerido que existe una severa multicolinealidad si $R_i^2 > .9$ para cualquier i. Algunos paquetes de software estadístico rechazarán incluir un predictor en el modelo cuando su valor R_i^2 sea muy cercano a 1.

No hay consenso entre los expertos en estadística en lo que respecta a qué soluciones son apropiadas cuando esté presente alguna severa multicolinealidad. Una posibilidad implica continuar con el uso de un modelo que incluya todos los predictores pero estimando parámetros con el uso de algo que no sean mínimos cuadrados. Para más detalles, consulte una referencia del capítulo.

EJERCICIOS Sección 13.5 (55–64)

55. El artículo "Bank Full Discharge of Rivers" (*Water Resources J.* 1978, 1141–1154) informa de datos acerca de la cantidad de descarga (q, en m^3/s), área de flujo (a, en m^2), y pendiente de la superficie del agua (b, en m/m) obtenidos en diversas estaciones del área de inundación. A continuación aparece un subconjunto de los datos. El artículo propuso un modelo multiplicativo de potencia $Q = \alpha a^{\beta} b^{\gamma} \epsilon$.

q	17.6	23.8	5.7	3.0	7.5
a	8.4	31.6	5.7	1.0	3.3
b	.0048	.0073	.0037	.0412	.0416
q	89.2	60.9	27.5	13.2	12.2
a	41.1	26.2	16.4	6.7	9.7
b	.0063	.0061	.0036	.0039	.0025

a. Utilice una transformación apropiada para hacer que el modelo sea lineal y luego estime los parámetros de regresión para el modelo transformado. Por último, estime α, β y γ (los parámetros del modelo original). ¿Cuál sería su predicción de cantidad de descarga cuando el área de flujo sea 10 y la pendiente sea de .01?

b. Sin hacer en realidad ningún análisis, ¿cómo ajustaría usted un modelo exponencial multiplicativo $Q = \alpha e^{\beta a} e^{\gamma b} \epsilon$?

c. Después de una transformación a linealidad en el inciso (a), un intervalo de confianza de 95% para el valor de la función de regresión transformada, cuando $a = 3.3$ y $b = .0046$, se obtuvo de la salida de computadora como (.217, 1.755). Obtenga un intervalo de confianza de 95% para $\alpha a^{\beta} b^{\gamma}$ cuando $a = 3.3$ y $b = .0046$.

56. En un experimento para estudiar factores que influyen en la gravedad específica de la madera ("Anatomical Factors Influencing Wood Specific Gravity of Slash Pines and the Implications for the Development of a High-Quality Pulpwood", *TAPPI*, 1964: 401-404), se obtuvo una muestra de 20 muestras de madera madura, y se tomaron medidas en el número de fibras/mm^2 en albura de primavera (x_1), número de fibras/mm^2 en albura de verano (x_2), % de albura de primavera (x_3), absorción de luz en albura de primavera (x_4), y absorción de luz en albura de verano (x_5).

a. El ajuste de la función de regresión $\mu_{Y \cdot x_1, x_2, x_3, x_4, x_5} = \beta_0 + \beta_1 x_1 + \cdots + \beta_5 x_5$ dio por resultado $R^2 = .769$. ¿Indican los datos que hay una relación lineal entre gravedad específica y al menos uno de los predictores? Pruebe usando $\alpha = .01$.

b. Cuando x_2 se elimina del modelo, el valor de R^2 permanece en .769. Calcule una R^2 ajustada para el modelo completo y el modelo con x_2 eliminada.

c. Cuando x_1, x_2 y x_4 se eliminan, el valor resultante de R^2 es .654. La suma total de cuadrados es SST = .0196610. ¿Sugieren los datos que x_1, x_2 y x_4 tienen coeficientes cero en el modelo de regresión verdadero? Pruebe las hipótesis relevantes al nivel .05.

d. La media y desviación estándar de x_3 fueron 52.540 y 5.4447, respectivamente, mientras que las de x_5 fueron 89.195 y 3.6660, respectivamente. Cuando se ajustó el modelo que comprende estas dos variables estandarizadas, la ecuación de regresión estimada fue $y = .5255 - .0236x_3' + .0097x_5'$. ¿Qué valor de gravedad específica pronosticaría el lector para una muestra de madera con % de albura de primavera = 50 y % de absorción de luz en albura de verano = 90?

e. La desviación estándar estimada del coeficiente estimado $\hat{\beta}_3$ de x_3' (es decir, para $\hat{\beta}_3$ del modelo estandarizado) fue .0046. Obtenga un intervalo de confianza de 95% para β_3.

f. Usando la información de los incisos (d) y (e), ¿cuál es el coeficiente estimado de x_3 en el modelo no estandarizado (usando sólo predictores x_3 y x_5), y cuál es la desviación estándar estimada del estimador de coeficiente (es decir, $s_{\hat{\beta}_3}$ para $\hat{\beta}_3$ en el modelo no estandarizado)?

g. La estimación de σ para el modelo de dos predictores es $s = .02001$, mientras que la desviación estándar estimada de $\hat{\beta}_0 + \hat{\beta}_3 x_3' + \hat{\beta}_5 x_5'$, cuando $x_3' = -.3747$ y $x_5' = -.2769$ (es decir, cuando $x_3 = 50.5$ y $x_5 = 88.9$) es .00482. Calcule un intervalo de predicción de 95% para gravedad específica cuando el % de albura de primavera = 50.5 y el % de absorción de luz en albura de verano = 88.9.

57. En la tabla siguiente, se presenta la SSE más pequeña para cada número de predictores k ($k = 1, 2, 3, 4$) para un problema de regresión en el que y = calor acumulativo de endurecimiento en cemento, x_1 = % de aluminato tricálcico, x_2 = % de silicato tricálcico, x_3 = % de ferrato de aluminio, y x_4 = % de silicato dicálcico.

Número de predictores k	Predictor(es)	SSE
1	x_4	880.85
2	x_1, x_2	58.01
3	x_1, x_2, x_3	49.20
4	x_1, x_2, x_3, x_4	47.86

Además, $n = 13$, y SST = 2715.76.

a. Use los criterios estudiados en el texto para recomendar el uso de un modelo particular de regresión.

b. ¿La selección directa produciría el mejor modelo de dos predictores? Explique.

58. El artículo "Response Surface Methodology for Protein Extraction Optimization of Red Pepper Seed" (*Food Sci. and*

Salida de Minitab para el ejercicio 58

Variables	R-cuadrado	R-cuadrado(adj)	Rojizas Cp	S	1 2 3 4 x x x x x x s s s s 1 1 1 2 2 3 x x x x q q q q x x x x x x 1 2 3 4 d d d d 2 3 4 3 4 4
1	52.7	50.9	174.4	0.73030	X
2	67.9	65.4	112.5	0.61349	X X
3	77.7	75.0	73.1	0.52124	X X X
4	83.4	80.7	50.8	0.45835	X X X X
5	90.9	88.9	21.4	0.34731	X X X X X
6	94.6	93.1	7.9	0.27422	X X X X X X
7	95.8	94.4	4.7	0.24683	X X X X X X X
8	96.2	94.6	5.1	0.24137	X X X X X X X X
9	96.4	94.7	6.1	0.23962	X X X X X X X X
10	96.6	94.6	7.5	0.24132	X X X X X X X X X
11	96.6	4.4	9.4	0.24716	X X X X X X X X X X
12	96.6	94.1	11.2	0.25328	X X X X X X X X X X
13	96.7	93.8	13.1	0.26041	X X X X X X X X X X X
14	96.7	93.4	15.0	0.26870	X X X X X X X X X X X X X X

Tech., 2010: 226–231) dio datos sobre la variable de respuesta y = proteína producida (%) y las variables independientes x_1 = temperatura (ºC), x_2 = pH, x_3 = tiempo de extracción (min) y x_4 = relación disolvente/alimento.

a. Ajustando el modelo con los cuatro x_i como predictores dio el siguiente resultado:

Predictor	Coef	DE Coef	T	P
Constante	−4.586	2.542	−1.80	0.084
x1	0.01317	0.02707	0.49	0.631
x2	1.6350	0.2707	6.04	0.000
x3	0.02883	0.01353	2.13	0.044
x4	0.05400	0.02707	1.99	0.058

Fuente	GL	SS	MS	F	P
Regresión	4	19.8882	4.9721	11.31	0.000
Error residual	24	10.5513	0.4396		
Total	28	30.4395			

Calcule e interprete los valores de R^2 y R^2 ajustado. ¿El modelo parece ser útil?

b. Ajustando el modelo completo de segundo orden, dio los siguientes resultados:

Predictor	Coef	DE Coef	T	P
Constante	−119.49	18.53	−6.45	0.000
x1	−0.1047	0.2839	−0.37	0.718
x2	28.678	3.625	7.91	0.000
x3	0.4074	0.1303	3.13	0.007
x4	0.2711	0.2606	1.04	0.316
x1sqd	−0.000752	0.002110	−0.36	0.727
x2sqd	−1.6452	0.2110	−7.80	0.000
x3sqd	0.0002121	0.0005275	0.40	0.694
x4sqd	−0.015152	0.002110	−7.18	0.000
x1x2	0.02150	0.02687	0.80	0.437
x1x3	0.000550	0.001344	0.41	0.688
x1x4	−0.000800	0.002687	−0.30	0.770
x2x3	−0.05900	0.01344	−4.39	0.001
x2x4	0.03900	0.02687	1.45	0.169
x3x4	0.002725	0.001344	2.03	0.062

S = 0.268703 R-cuadrado = 96.7% R-cuadrado(adj) = 93.4%

Fuente	GL	SS	MS	F	P
Regresión	14	29.4287	2.1020	29.11	0.000
Error residual	14	1.0108	0.0722		
Total	28	30.4395			

¿Por lo menos uno de los predictores de segundo orden parece ser útil? Lleve a cabo una prueba adecuada de la hipótesis.

c. Desde la salida en (b), una conjetura razonable es que ninguno de los predictores de la participación x_1 están proporcionando información útil. Cuando se eliminan estos predictores, el valor de SSE para el modelo de regresión se reduce 1.1887. ¿Éste apoya la conjetura?

d. Aquí está la salida de la mejor opción de subconjuntos Minitab, con sólo el subconjunto único y mejor de cada tamaño identificado. ¿Qué modelo(s) usted consideraría usar (sujeto a comprobar la adecuación del modelo)?

59. La opción de Mejor Regresión de Minitab se utilizó en los datos de gravedad específica de madera del ejercicio 56, produciendo la siguiente salida de computadora. ¿Cuál(es) modelo(s) recomendaría el lector que se investigara en más detalle?

Respuesta es gravedad específica

Variables	R-cuadrado	R-cuadrado (adj)	C-p	s	b fibra primavera	b fibra verano	s absorción de la luz primavera	s madera primavera	s absorción de la luz verano
1	56.4	53.9	10.6	0.021832			X		
1	10.6	5.7	38.5	0.031245					X
1	5.3	0.1	41.7	0.032155		X			
2	65.5	61.4	7.0	0.019975			X		X
2	62.1	57.6	9.1	0.020950		X	X		
2	60.3	55.6	10.2	0.021439	X		X		
3	72.3	67.1	4.9	0.018461	X		X		X
3	71.2	65.8	5.6	0.018807			X	X	X
3	71.1	65.7	5.6	0.018846		X	X		X
4	77.0	70.9	4.0	0.017353	X		X	X	X
4	74.8	68.1	5.4	0.018179		X	X	X	X
4	72.7	65.4	6.7	0.018919	X	X	X		X
5	77.0	68.9	6.0	0.017953	X	X	X	X	X

60. La siguiente salida impresa de Minitab resultó de aplicar el método de eliminación inversa, y el método de selección directa, a los datos de gravedad específica de madera de que trata el ejercicio 56. Para cada uno de los métodos, explique qué ocurrió en cada iteración del algoritmo.

```
La respuesta es gravedad específica en 5
predictores, con N = 20

Paso            1        2        3        4
Constante   0.4421   0.4384   0.4381   0.5179

fibra
primavera  0.00011  0.00011  0.00012
Valor T       1.17     1.95     1.98

fibra
verano     0.00001
Valor T       0.12

%madera
primavera −0.00531  −0.00526  −0.00498  −0.00438
Valor T      −5.70     −6.56     −5.96     −5.20

absorción
de luz
primavera  −0.0018   −0.0019
Valor T      −1.63     −1.76

absorción
de luz
verano      0.0044    0.0044    0.0031    0.0027
Valor T       3.01      3.31      2.63      2.12

S           0.0180    0.0174    0.0185    0.0200
R-cuadrado   77.05     77.03     72.27     65.50

Paso            1         2
Constante   0.7585    0.5179

%madera
primavera −0.00444  −0.00438
Valor T      −4.82     −5.20
```

(*continúa en la siguiente columna*)

```
absorción
de luz
verano                0.0027
Valor T                 2.12

S           0.0218    0.0200
R-cuadrado   56.36     65.50
```

61. Reconsidere los datos de gravedad específica de madera de que habla el ejercicio 56. Los siguientes valores de R^2 resultaron de hacer regresión en cada predictor en los otros cuatro predictores (en la primera regresión, la variable dependiente era x_1 y los predictores fueron x_2-x_5, etc.): .628, .711, .341, .403, y .403. ¿La multicolinealidad parece ser un problema importante? Explique.

62. Un estudio realizado para investigar la relación entre una variable de respuesta, que relaciona caídas de presión en una columna de burbujas de placa de filtros y los predictores x_1 = velocidad superficial del fluido, x_2 = viscosidad del líquido, y x_3 = medida de mallas, produjo los datos siguientes ("A Correlation of Two-Phase Pressure Drops in Screen-Plate Bubble Column", *Canad. J. of Chem. Engr.*, 1993: 460–463). Los residuos estandarizados y valores h_{ii} resultaron del modelo con sólo x_1, x_2, y x_3 como predictores. ¿Hay algunas observaciones poco comunes?

63. La salida de regresión múltiple de Minitab, para los datos de hidrocarburos poliaromáticos (PAH) del ejercicio 53 de la sección anterior incluyó la información siguiente:

```
Observaciones inusuales

          fluor-
Obs    x1  etano Ajuste DE Ajuste Residual St Resid
  6 243500  604.7 582.9   40.7     21.8    1.25X
  7  67793   27.7 139.3   12.3   −111.6   −2.62R

R denota una observación con un gran residuo
estandarizado

X denota una observación cuyo valor X tiene gran
influencia.
```

Datos para el ejercicio 62

Observación	Velocidad	Viscosidad	Medida de mallas	Respuesta	Residuo estandarizado	h_{ii}
1	2.14	10.00	.34	28.9	2.01721	.202242
2	4.14	10.00	.34	26.1	1.34706	.066929
3	8.15	10.00	.34	22.8	.96537	.274393
4	2.14	2.63	.34	24.2	1.29177	.224518
5	4.14	2.63	.34	15.7	−.68311	.079651
6	8.15	2.63	.34	18.3	.23785	.267959
7	5.60	1.25	.34	18.1	.06456	.076001
8	4.30	2.63	.34	19.1	.13131	.074927
9	4.30	2.63	.34	15.4	−.74091	.074927
10	5.60	10.10	.25	12.0	−1.38857	.152317
11	5.60	10.10	.34	19.8	−.03585	.068468
12	4.30	10.10	.34	18.6	−.40699	.062849
13	2.40	10.10	.34	13.2	−1.92274	.175421
14	5.60	10.10	.55	22.8	−1.07990	.712933
15	2.14	112.00	.34	41.8	−1.19311	.516298
16	4.14	112.00	.34	48.6	1.21302	.513214
17	5.60	10.10	.25	19.2	.38451	.152317
18	5.60	10.10	.25	18.4	.18750	.152317
19	5.60	10.10	.25	15.0	−.64979	.152317

¿Qué sugiere esto acerca de lo apropiado de usar la ecuación ajustada previamente dada como base para inferencias? Los investigadores en realidad eliminaron la observación #7 e hicieron de nuevo una regresión. ¿Tiene sentido esto?

64. Consulte los datos de descarga de agua dados en el ejercicio 55 y haga $y = \ln(q)$, $x_1 = \ln(a)$, y $x_2 = \ln(b)$. Considere ajustar el modelo $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$.

a. Las h_{ii} resultantes son .138, .302, .266, .604, .464, .360, .215, .153, .214 y .284. ¿Alguna de estas observaciones parece ser influyente?

b. Los coeficientes estimados son $\hat{\beta}_0 = 1.5652$, $\hat{\beta}_1 = .9450$, y $\hat{\beta}_2 = .1815$, y las correspondientes desviaciones estándar estimadas son $s_{\hat{\beta}_0} = .7328$, $s_{\hat{\beta}_1} = .1528$ y $s_{\hat{\beta}_2} = .1752$. El segundo residuo estandarizado es $e_2^* = 2.19$. Cuando del conjunto de datos se omite la segunda observación, los coeficientes estimados resultantes son $\hat{\beta}_0 = 1.8982$, $\hat{\beta}_1 = 1.025$ y $\hat{\beta}_2 = .3085$. ¿Alguno de estos cambios indica que la segunda observación es influyente?

c. La eliminación de la cuarta observación (¿por qué?) da $\hat{\beta}_0 = 1.4592$, $\hat{\beta}_1 = .9850$ y $\hat{\beta}_2 = .1515$. ¿Es influyente esta observación?

EJERCICIOS SUPLEMENTARIOS (65–82)

65. Se sabe que curar el concreto es vulnerable a vibraciones de choque, que pueden causar agrietamiento o daños ocultos al material. Como parte de un estudio de fenómenos de vibración, el artículo "Shock Vibration Test of Concrete" (*ACI Materials J.*, 2002: 361–370) informó de los datos siguientes acerca de la velocidad máxima de una partícula (mm/s) y la relación entre la velocidad ultrasónica de un pulso después del choque y la velocidad antes del impacto en prismas de concreto.

Obs	vmp	Relación	Obs	vmp	Relación
1	160	.996	16	708	.990
2	164	.996	17	806	.984
3	178	.999	18	884	.986
4	252	.997	19	526	.991
5	293	.993	20	490	.993
6	289	.997	21	598	.993
7	415	.999	22	505	.993
8	478	.997	23	525	.990
9	391	.992	24	675	.991
10	486	.985	25	1211	.981
11	604	.995	26	1036	.986
12	528	.995	27	1000	.984
13	749	.994	28	1151	.982
14	772	.994	29	1144	.962
15	532	.987	30	1068	.986

Aparecieron grietas transversales en los últimos 12 prismas, mientras que no se observó agrietamiento en los primeros 18 prismas.

a. Construya una gráfica de caja comparativa de la velocidad máxima de partículas (vmp) para los prismas agrietados y no agrietados y comente. A continuación estime la diferencia entre la vmp promedio verdadera para prismas agrietados y no agrietados en una forma que exprese información acerca de la precisión y confiabilidad.

b. Los investigadores ajustaron el modelo de regresión lineal simple a todo el conjunto de datos formado por 30 observaciones, con la vmp como variable independiente y la relación como la variable dependiente. Utilice un paquete de software de estadística para ajustar varios modelos de regresión diferentes, y saque inferencias apropiadas.

66. Los autores del artículo "Long-Term Effects of Cathodic Protection on Prestressed Concrete Structures" (*Corrosion*, 1997: 891–908) presentó un diagrama de dispersión de $y =$ flujo de permeabilidad en estado estable (μA/cm^2) en función de $x =$ grosor inverso de hoja metálica (cm^{-1}); el patrón lineal sustancial se usó como base para una importante conclusión acerca del comportamiento del material. A continuación aparece la salida de Minitab del ajuste del modelo de regresión lineal simple a los datos.

```
La ecuación de regresión es
flujo = −0.398 + 0.260 grosor inverso

Predictor        Coef         DE t-cociente        P
Constante    −0.3982       0.5051      −0.79    0.460
grosor
inverso      0.26042    0.01502       17.34    0.000

s = 0.4506 R-cuadrado = 98.0% R-cuadrado(adj) = 97.7%

Análisis de varianza

Fuente        GL        SS        MS         F      P
Regresión      1    61.050    61.050    300.64  0.000
Error          6     1.218     0.203
Total          7    62.269

     grosor-                        DE.           St.
Obs. inverso  flujo   Ajuste  Ajuste  Residual  Resid
  1    19.8    4.3    4.758   0.242    −0.458  −1.20
  2    20.6    5.6    4.966   0.233     0.634   1.64
  3    23.5    6.1    5.722   0.203     0.378   0.94
  4    26.1    6.2    6.399   0.182    −0.199  −0.48
  5    30.3    6.9    7.493   0.161    −0.593  −1.41
  6    43.5   11.2   10.930   0.236     0.270   0.70
  7    45.0   11.3   11.321   0.253    −0.021  −0.06
  8    46.5   11.7   11.711   0.271    −0.011  −0.03
```

a. Interprete la pendiente estimada y el coeficiente de determinación.
b. Calcule una estimación puntual del flujo promedio verdadero cuando el grosor inverso de la hoja metálica sea de 23.5.
c. ¿Parece útil el modelo?
d. Pronostique el flujo cuando el grosor inverso es 45 de modo que exprese información acerca de la precisión y la confiabilidad.
e. Investigue lo adecuado del modelo.

67. El artículo "Validation of the Rockport Fitness Walking Test in College Males and Females" (*Research Quarterly for Exercise and Sport*, 1994: 152–158) recomendó la siguiente ecuación de regresión estimada para relacionar $y = VO_2$máx (L/min, una medida de la salud cardiorrespiratoria) con los predictores $x_1 =$ género (femenino = 0, masculino = 1), $x_2 =$ peso (lb), $x_3 =$ tiempo para recorrer 1 milla (min), y $x_4 =$ ritmo cardiaco al final de la caminata (pulsaciones /min)

$$y = 3.5959 + .6566x_1 + .0096x_2 - .0996x_3 - .0080x_4$$

a. ¿Cómo interpretaría el coeficiente estimado $\hat{\beta}_3 = -.0996$?
b. ¿Cómo interpretaría el coeficiente estimado $\hat{\beta}_1 = .6566$?
c. Suponga que los datos de un hombre cuyo peso fue de 170 lb, con tiempo de caminata de 11 minutos, y ritmo cardiaco de 140 pulsaciones/min, produjo un VO_2máx = 3.15. ¿Qué pronosticaría el lector para VO_2máx en esta situación y cuál es el valor del residuo correspondiente?
d. Usando SSE = 30.1033 y SST = 102.3922, ¿qué proporción de variación observada en VO_2máx se puede atribuir a la relación con el modelo?
e. Suponiendo un tamaño muestral de $n = 20$, realice una prueba de las hipótesis para determinar si el modelo seleccionado especifica una relación útil entre VO_2máx y al menos uno de los predictores.

68. El reconocimiento de características de modelos de superficie de piezas complicadas se está haciendo cada vez más importante en el desarrollo de eficientes sistemas de diseño asistido por computadora (CAD). El artículo "A Computationally Efficient Approach to Feature Abstraction in Design-Manufacturing Integration" (*J. of Engr. for Industry*, 1995: 16–27) contenía una gráfica de $\log_{10}$(tiempo total de reconocimiento), con tiempo en segundos, en función de $\log_{10}$(número de aristas de una pieza), de la cual se leyeron los siguientes valores representativos:

Log(aristas)	1.1	1.5	1.7	1.9	2.0	2.1
Log(tiempo)	.30	.50	.55	.52	.85	.98
Log(aristas)	2.2	2.3	2.7	2.8	3.0	3.3
Log(tiempo)	1.10	1.00	1.18	1.45	1.65	1.84
Log(aristas)	3.5	3.8	4.2	4.3		
Log(tiempo)	2.05	2.46	2.50	2.76		

a. ¿Un diagrama de dispersión de log(tiempo) en función de log(aristas) sugiere una relación lineal aproximada entre estas dos variables?
b. ¿Qué modelo probabilístico para relacionar $y =$ tiempo de reconocimiento con $x =$ número de aristas está implicado por la relación de regresión lineal simple entre las variables transformadas?
c. Las cantidades calculadas en resumen a partir de los datos son

$n = 16 \quad \sum x_i' = 42.4 \quad \sum y_i' = 21.69$
$\sum (x_i')^2 = 126.34 \quad \sum (y_i')^2 = 38.5305$
$\sum x_i' y_i' = 68.640$

Calcule estimaciones de los parámetros para el modelo del inciso (b), y a continuación obtenga una predicción puntual del tiempo cuando el número de aristas sea de 300.

69. La presión de aire (libras por pulgada cuadrada) y la temperatura (°F) se midieron para un proceso de compresión de cierto aparato de émbolo y cilindro y produjeron los datos siguientes (de *Introduction to Engineering Experimentation*, Prentice-Hall, Inc., 1996, p. 153):

Presión	20.0	40.4	60.8	80.2	100.4
Temperatura	44.9	102.4	142.3	164.8	192.2
Presión	120.3	141.1	161.4	181.9	201.4
Temperatura	221.4	228.4	249.5	269.4	270.8
Presión	220.8	241.8	261.1	280.4	300.1
Temperatura	291.5	287.3	313.3	322.3	325.8
Presión	320.6	341.1	360.8		
Temperatura	337.0	332.6	342.9		

a. ¿Ajustaría el lector el modelo de regresión lineal simple a los datos, y lo usaría como base para pronosticar la temperatura a partir de la presión? ¿Por qué sí o por qué no?
b. Encuentre un modelo probabilístico apropiado y, del modo más informativo posible, úselo como base para predecir el valor de temperatura que resultaría de una presión de 200.

70. Un estudiante de ingeniería aeronáutica realizó un experimento, para estudiar en qué forma la relación $y =$ sustentación/resistencia al avance, relacionada con las variables $x_1 =$ posición de cierta superficie elevadora hacia adelante respecto al ala principal, y $x_2 =$ posición de la cola con respecto al ala principal; obtuvo los datos siguientes (*Statistics for Engineering Problem Solving*, p. 133):

x (pulgadas)	x (pulgadas)	y
−1.2	−1.2	.858
−1.2	0	3.156
−1.2	1.2	3.644
0	−1.2	4.281
0	0	3.481
0	1.2	3.918
1.2	−1.2	4.136
1.2	0	3.364
1.2	1.2	4.018

$\bar{y} = 3.428$, SST = 8.55

a. El ajuste del modelo de primer orden da SSE = 5.18, mientras que incluir $x_3 = x_1x_2$ como predictor produce SSE = 3.07. Calcule e interprete el coeficiente de determinación múltiple para cada modelo.

b. Efectúe una prueba de utilidad de modelo usando $\alpha = .05$ para cada uno de los modelos descritos en el inciso (a). ¿Le sorprende cualquiera de los dos resultados?

71. Un baño de amoniaco es el más utilizado para depositar capas de aleación de Pd-Ni. El artículo "Modelling of Palladium and Nickel in an Ammonia Bath in a Rotary Device" (*Plating and Surface Finishing*, 1997: 102–104) informó de una investigación sobre la forma en que las características de la composición del baño afectan las propiedades de la capa. Tenga en cuenta los siguientes datos en x_1 = concentración de Pd (g/dm^3), x_2 = concentración de Ni (g/dm^3), x_3 = pH, x_4 = temperatura (°C), x_5 = densidad de corriente catódica (A/dm^2) y y = contenido de paladio (%) de la capa.

Obs	concpd	concni	pH	temp	denscorr	contpad
1	16	24	9.0	35	5	61.5
2	8	24	9.0	35	3	51.0
3	16	16	9.0	35	3	81.0
4	8	16	9.0	35	5	50.9
5	16	24	8.0	35	3	66.7
6	8	24	8.0	35	5	48.8
7	16	16	8.0	35	5	71.3
8	8	16	8.0	35	3	62.8
9	16	24	9.0	25	3	64.0
10	8	24	9.0	25	5	37.7
11	16	16	9.0	25	5	68.7
12	8	16	9.0	25	3	54.1
13	16	24	8.0	25	5	61.6
14	8	24	8.0	25	3	48.0
15	16	16	8.0	25	3	73.2
16	8	16	8.0	25	5	43.3
17	4	20	8.5	30	4	35.0
18	20	20	8.5	30	4	69.6
19	12	12	8.5	30	4	70.0
20	12	28	8.5	30	4	48.2
21	12	20	7.5	30	4	56.0
22	12	20	9.5	30	4	77.6
23	12	20	8.5	20	4	55.0
24	12	20	8.5	40	4	60.6
25	12	20	8.5	30	2	54.9
26	12	20	8.5	30	6	49.8
27	12	20	8.5	30	4	54.1
28	12	20	8.5	30	4	61.2
29	12	20	8.5	30	4	52.5
30	12	20	8.5	30	4	57.1
31	12	20	8.5	30	4	52.5
32	12	20	8.5	30	4	56.6

a. Ajuste el modelo de primer orden con los cinco predictores y evalúe su utilidad. ¿Todos los predictores parecen importantes?

b. Ajuste el modelo completo de segundo orden y evalúe su utilidad.

c. El grupo de predictores de segundo orden (de interacción y cuadráticos), ¿parece dar más información útil acerca de y que aquella con la que contribuyen los predictores de primer orden? Realice una prueba apropiada de las hipótesis.

d. Los autores del artículo citado recomendaron el uso de los cinco predictores de primer orden más el predictor adicional $x_6 = (\text{pH})^2$. Ajuste este modelo. ¿Los seis predictores parecen importantes?

72. El artículo "An Experimental Study of Resistance Spot Welding in 1 mm Thick Sheet of Low Carbon Steel" (*J. of Engr. Manufacture*, 1996: 341–348) examinó un análisis estadístico cuyo objetivo básico era establecer una relación que pudiera explicar la variación en resistencia de soldaduras (y), al relacionar la resistencia con las características del proceso como son corriente de soldadura (wc), tiempo de soldadura (wt) y fuerza del electrodo (ef).

a. SST = 16.18555 y el ajuste del modelo completo de segundo orden dio SSE = .80017. Calcule e interprete el coeficiente de determinación múltiple.

b. Suponiendo que $n = 37$, efectúe una prueba de utilidad del modelo (la tabla ANOVA del artículo indica que $n - (k + 1) = 1$, pero otra información dada contradice esto y es consistente con el tamaño muestral sugerido).

c. La relación F dada para la interacción entre corriente y tiempo fue de 2.32. Si todos los predictores se retienen en el modelo, ¿puede eliminarse este predictor de interacción? [*Sugerencia:* al igual que en regresión lineal simple, una relación F para un coeficiente es el cuadrado de su razón t.]

d. Los autores propusieron eliminar dos predictores de interacción y un predictor cuadrático y recomendaron la ecuación estimada $y = 3.352 + .098\text{wc} + .222\text{wt} + .297\text{ef} - .0102(\text{wt})^2 - .037(\text{ef})^2 + .0128(\text{wc})(\text{wt})$. Considere una corriente de soldadura de 10 kA, un tiempo de soldadura de 12 ciclos de ca y una fuerza de electrodo de 6 kN. Suponiendo que la desviación estándar estimada de la resistencia pronosticada en esta situación sea .0750, calcule un intervalo de predicción de 95% para la resistencia. ¿El intervalo sugiere que el valor de la resistencia pueda pronosticarse con precisión?

73. La información siguiente sobre x = frecuencia (MHz) y y = potencia de salida (W), para cierto tipo de configuración láser, apareció en una gráfica del artículo "Frequency Dependence in RF Discharge Excited Waveguide CO_2 Lasers" (*IEEE J. of Quantum Electronics*, 1984: 509–514).

x	60	63	77	100	125	157	186	222
y	16	17	19	21	22	20	15	5

Un análisis computarizado dio la siguiente información para un modelo de regresión cuadrático: $\hat{\beta}_0 = -1.5127$, $\hat{\beta}_1 = .391901$, $\hat{\beta}_2 = -.00163141$, $s_{\hat{\beta}_2} = .00003391$, SSE = .29, SST = 202.88 y $s_{\hat{Y}} = .1141$ cuando $x = 100$.

a. ¿El modelo cuadrático parece apropiado para explicar la variación observada en potencia de salida al relacionarla con la frecuencia?

b. El modelo de regresión lineal simple ¿sería casi tan satisfactorio como el modelo cuadrático?

c. ¿Piensa usted que valdría la pena considerar un modelo cúbico?

d. Calcule un intervalo de confianza de 95% para salida esperada de potencia cuando la frecuencia es de 100.

e. Use un intervalo de predicción de 95% para predecir la potencia desde una sola prueba experimental cuando la frecuencia es de 100.

74. La conductividad es una importante característica del vidrio. El artículo "Structure and Properties of Rapidly Quenched Li_2O-Al_2O-

Nb_2O_5 Glasses" (*J. of the Amer. Ceramic Soc.*, 1983: 890–892) informa de los datos siguientes acerca del contenido de $x = Li_2O$ de cierto tipo de vidrio y y = conductividad a 500 K.

x	19	20	24	27	29	30
y	$10^{-8.0}$	$10^{-7.1}$	$10^{-7.2}$	$10^{-6.7}$	$10^{-6.2}$	$10^{-6.8}$
x	31	39	40	43	45	50
y	$10^{-5.8}$	$10^{-5.3}$	$10^{-6.0}$	$10^{-4.7}$	$10^{-5.4}$	$10^{-5.1}$

(Éste es un subconjunto de los datos que aparecieron en el artículo.) Proponga un modelo apropiado para relacionar y con x, estimar los parámetros del modelo, y predecir la conductividad cuando el contenido de Li_2O es de 35.

75. El efecto del manganeso (Mn) en el crecimiento del trigo se examina en el artículo "Manganese Deficiency and Toxicity Effects on Growth, Development and Nutrient Composition in Wheat" (*Agronomy J.*, 1984: 213-217). Se utilizó un modelo de regresión cuadrático para relacionar y = altura de la planta (cm) con $x = \log_{10}$(Mn agregado), con μM como las unidades para el Mn agregado. La información siguiente apareció en un diagrama de dispersión del artículo.

x	−1.0	−.4	0	.2	1.0
y	32	37	44	45	46
x	2.0	2.8	3.2	3.4	4.0
y	42	42	40	37	30

Además, $\hat{\beta}_0 = 41.7422$, $\hat{\beta}_1 = 6.581$, $\hat{\beta}_2 = -2.3621$, $s_{\hat{\beta}_0} = .8522$, $s_{\hat{\beta}_1} = 1.002$, $s_{\hat{\beta}_2} = .3073$ y SSE = 26.98.

a. ¿Es útil el modelo cuadrático para describir la relación entre x y y? [*Sugerencia:* la regresión cuadrática es un caso especial de regresión múltiple con $k = 2$, $x_1 = x$, y $x_2 = x^2$.] Aplique un procedimiento apropiado.

b. ¿Debe eliminarse el predictor cuadrático?

c. Estime la altura esperada para el trigo tratado con 10 μM de Mn usando un intervalo de confianza de 90%. [*Sugerencia:* la desviación estándar estimada de $\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2$ es 1.031].

76. El artículo "Chemithermomechanical Pulp from Mixed High Density Hardwoods" (*TAPPI*, julio de 1988: 145–146) informa de un estudio en el que se obtuvo la información siguiente para relacionar y = área superficial específica (cm^2/g) con x_1 = % de NaOH utilizado como sustancia química de tratamiento previo y x_2 = tiempo de tratamiento (min) para un lote de pulpa.

x_1	x_2	y
3	30	5.95
3	60	5.60
3	90	5.44
9	30	6.22
9	60	5.85
9	90	5.61
15	30	8.36
15	60	7.30
15	90	6.43

La siguiente salida de Minitab resultó de una petición para ajustar el modelo $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$.

```
La ecuación de regresión es
AREA = 6.05 + 0.142 NAOH − 0.0169 TIEMPO

Predictor      Coef         DE   cociente-t        p
Constante    6.0483     0.5208        11.61    0.000
NAOH        0.14167    0.03301         4.29    0.005
TIEMPO    −0.016944   0.006601        −2.57    0.043

S = 0.4851  R-cuadrado = 80.7%  R-cuadrado(adj) = 74.2%

Análisis de varianza

FUENTE       DF       SS       MS       F       p
Regresión     2   5.8854   2.9427   12.51   0.007
Error         6   1.4118   0.2353
Total         8   7.2972
```

a. ¿Qué proporción de variación observada en área superficial específica puede ser explicada por la relación del modelo?

b. ¿El modelo seleccionado parece especificar una relación útil entre la variable dependiente y los predictores?

c. Siempre que el % de NaOH permanezca en el modelo, ¿sugeriría usted que se eliminara el *tiempo de tratamiento* del predictor?

d. Calcule un intervalo de confianza de 95% para el cambio esperado en el área superficial específica asociado con un aumento de 1% en NaOH, cuando el tiempo de tratamiento se mantiene fijo.

e. Minitab reportó que la desviación estándar estimada de $\hat{\beta}_0 + \hat{\beta}_1(9) + \hat{\beta}_2(60)$ es .162. Calcule un intervalo de predicción para el valor del área superficial específica a observar cuando el % de NaOH = 9 y el tiempo de tratamiento = 60.

77. El artículo "Sensitivity Analysis of a 2.5 kW Proton Exchange Membrane Fuel Cell Stack by Statistical Method" (*J. of Fuel Cell Sci. and Tech.*, 2009: 1–6) utilizó análisis de regresión para investigar la relación entre la potencia de la celda de combustible (W) y las variables independientes x_1 = presión de H_2 (psi), x_2 = flujo de H_2 (stoc), x_3 = presión de aire (psi) y x_4 = flujo de aire (stoc).

a. En seguida se muestra la salida de Minitab de ajuste del modelo con las variables antes mencionadas como predictores independientes (también se ajustan por los autores del artículo citado):

```
Predictor      Coef    SE Coef        T        P
Constante    1507.3      206.8     7.29    0.000
x1           −4.282      4.969    −0.86    0.407
x2             7.46      62.11     0.12    0.907
x3          −0.9162     0.6227    −1.47    0.169
x4            90.60      24.84     3.65    0.004

S = 4.6885  R-cuadrado = 59.6%  R-cuadrado(adj) = 44.9%

Fuente           DF      SS      MS      F      P
Regresión         4   40048   10012   4.06  0.029
Error residual   11   27158    2469
Total            15   67206
```

a. ¿Parece haber una relación útil entre la potencia y por lo menos uno de los indicadores? Lleve a cabo una prueba formal de hipótesis.

b. Ajustar el modelo con predictores x_3, x_4, y la interacción x_3x_4 dio $R^2 = .834$. ¿Este modelo parece ser útil? ¿Puede utili-

zarse una prueba de F para comparar este modelo con el de (a)? Explique.

c. Ajustar el modelo con predictores x_1– x_4, así como todas las interacciones de segundo orden dio $R^2 = .960$ (este modelo también se ajusta por los investigadores). ¿Parece que al menos uno de los predictores de la interacción proporciona información útil sobre la potencia más allá de lo previsto por los predictores de primer orden? Establezca y ponga a prueba las hipótesis adecuadas con un nivel de significancia de .05.

78. La fibra de coco, derivada del coco, es un material respetuoso del medio ambiente con un gran potencial para su uso en la construcción. El artículo "Seepage Velocity and Piping Resistance of Coir Fiber Mixed Soils" (*J. of Irrig. and Drainage Engr.*, 2008: 485–492) incluye varios análisis de regresión múltiple. Los autores del artículo amablemente proporcionaron los datos adjuntos x_1 = el contenido de fibra (%), x_2 = longitud de la fibra (mm), x_3 = gradiente hidráulico (sin unidades), y y = velocidad de infiltración (cm/s).

Obs	contenido	longitud	gradiente	velocidad
1	0.0	0	0.400	0.027
2	0.0	0	0.716	0.050
3	0.0	0	0.925	0.080
4	0.0	0	1.098	0.099
5	0.0	0	1.226	0.107
6	0.0	0	1.427	0.140
7	0.0	0	1.709	0.178
8	0.0	0	1.872	0.200
9	0.5	50	0.380	0.022
10	0.5	50	0.774	0.040
11	0.5	50	1.056	0.060
12	0.5	50	1.329	0.111
13	0.5	50	1.598	0.158
14	0.5	50	1.799	0.188
15	1.0	50	0.410	0.026
16	1.0	50	0.577	0.038
17	1.0	50	0.748	0.049
18	1.0	50	0.927	0.060
19	1.0	50	1.090	0.070
20	1.0	50	1.239	0.088
21	1.0	50	1.496	0.111
22	1.0	50	1.744	0.134
23	1.0	50	1.915	0.145
24	1.5	50	0.444	0.014
25	1.5	50	0.821	0.037
26	1.5	50	1.141	0.058
27	1.5	50	1.474	0.082
28	1.5	50	1.581	0.112
29	1.5	50	1.983	0.144
30	1.0	25	0.462	0.028
31	1.0	25	0.705	0.059
32	1.0	25	0.987	0.084
33	1.0	25	1.154	0.101
34	1.0	25	1.479	0.150
35	1.0	25	1.786	0.194
36	1.0	25	1.957	0.218
37	1.0	40	0.419	0.030
38	1.0	40	0.705	0.050
39	1.0	40	0.979	0.068
40	1.0	40	1.226	0.091
41	1.0	40	1.470	0.126
42	1.0	40	1.744	0.168
43	1.0	60	0.436	0.034
44	1.0	60	0.650	0.051
45	1.0	60	0.889	0.068
46	1.0	60	1.222	0.093

(*continúa en la siguiente columna*)

Obs	contenido	longitud	gradiente	velocidad
47	1.0	60	1.477	0.112
48	1.0	60	1.726	0.139
49	1.0	60	1.983	0.173

a. La siguiente es la salida de ajuste del modelo con los tres x_i como predictores:

Predictor	Coef	SE Coef	T	P
Constante	−0.002997	0.007639	−0.39	0.697
contenido de fibra	−0.012125	0.007454	−1.63	0.111
longitud de fibra	−0.0003020	0.0001676	−1.80	0.078
gradiente hidráulico	0.102489	0.004711	21.76	0.000

S = 0.0162355 R-Cuadrado = 91.6% R- R-Cuadrado(adj) = 91.1%

Fuente	GL	SS	MS	F	P
Regresión	3	0.129898	0.043299	164.27	0.000
Error residual	45	0.011862	0.000264		
Total	48	0.141760			

¿Cómo interpretaría el número –.0003020 en la columna Coef en la salida?

b. ¿El contenido de fibra parece proporcionar información útil acerca de la velocidad, siempre que la longitud de la fibra y el gradiente hidráulico permanezcan en el modelo? Lleve a cabo una prueba de hipótesis.

c. Ajustar el modelo con una longitud de fibra exacta y el gradiente hidráulico como predictores dio los coeficientes de regresión estimados $\hat{\beta}_0 = -.005315$, $\hat{\beta}_1 = -.0004968$ y $\hat{\beta}_2 = .102204$ (las relaciones de t para estas dos variables predictoras son altamente significativas). Además, $s_{\hat{Y}} = .00286$ cuando la longitud de fibra = 25 y el gradiente hidráulico = 1.2. ¿Hay pruebas convincentes de que la velocidad promedio real es algo más que .1 en esta situación? Lleve a cabo una prueba con un nivel de significancia de .05.

d. El ajuste del modelo completo de segundo orden (al igual que los autores del artículo) dio lugar a SSE = .003579. ¿Parece que al menos uno de los predictores de segundo orden proporciona información útil por encima de lo previsto por los tres predictores de primer orden? Pruebe las hipótesis pertinentes.

79. El artículo "A Statistical Analysis of the Notch Toughness of 9% Nickel Steels Obtained from Production Heats" (*J. of Testing and Eval.*, 1987: 355–363) informa de los resultados de un análisis de regresión múltiple que relaciona la resistencia Charpy de canal en v y (en joules) con las siguientes variables: x_1 = grosor de placa (mm), x_2 = contenido de carbono (%), x_3 = contenido de manganeso (%), x_4 = contenido de fósforo (%), x_5 = contenido de azufre (%), x_6 = contenido de silicio (%), x_7 = contenido de níquel (%), x_8 = punto de fluencia (Pa), y x_9 = resistencia a la tensión (Pa).

a. Los mejores subconjuntos posibles involucraron sumar variables en el orden x_5, x_8, x_6, x_3, x_2, x_7, x_9, x_1, y x_4. Los valores de R_k^2, MSE_k y C_k son como sigue:

Núm. de predictores	**1**	**2**	**3**	**4**
$\boldsymbol{R_k^2}$	.354	.453	.511	.550
$\textbf{MSE}_k$	2295	1948	1742	1607
$\boldsymbol{C_k}$	314	173	89.6	35.7

Núm. de predictores	5	6	7	8	9
R_k^2	.562	.570	.572	.575	.575
MSE_k	1566	1541	1535	1530	1532
C_k	19.9	11.0	9.4	8.2	10.0

¿Qué modelo recomendaría el lector? Explique la justificación de su elección.

b. Los autores también consideraron modelos de segundo orden que comprendían predictores x_j^2 y $x_i x_j$. la información sobre los mejores de estos modelos, comenzando con las variables x_2, x_3, x_5, x_6, x_7 y x_8 es como sigue (al pasar del mejor modelo de cuatro predictores al mejor modelo de cinco predictores, x_8 se eliminó y se introdujeron x_2x_6 y x_7x_8; x_8 se volvió a introducir en una etapa posterior):

Núm. de predictores	1	2	3	4	5
R_k^2	.415	.541	.600	.629	.650
MSE_k	2079	1636	1427	1324	1251
C_k	433	109	104	52.4	16.5

Núm. de predictores	6	7	8	9	10
R_k^2	.652	.655	.658	.659	.659
MSE_k	1246	1237	1229	1229	1230
C_k	14.9	11.2	8.5	9.2	11.0

¿Cuál de estos modelos recomendaría el lector, y por qué? [*Nota:* los modelos basados en ocho de las variables originales no dio una mejoría marcada sobre aquellas bajo consideración aquí.]

80. Se seleccionó una muestra de $n = 20$ compañías, y los valores de y = precio de acción y $k = 15$ variables (por ejemplo dividendos trimestrales, utilidades del año previo, y relación de deuda) se determinaron. Cuando el modelo de regresión múltiple que utilizó estos 15 predictores se ajustó a los datos, resultó $R^2 = .90$.

a. ¿El modelo parece especificar una relación útil entre y y las variables predictoras? Efectúe una prueba usando nivel de significancia .05. [*Sugerencia:* el valor crítico F para grado de libertad de numerador 15 y denominador 4 es 5.86.]

b. Con base en el resultado del inciso (a), ¿el valor de R^2 implica por sí mismo que un modelo es útil? ¿Bajo qué circunstancias podría sospecharse de un modelo con un alto valor de R^2?

c. Con n y k como se dan previamente, ¿qué tan grande tendría que ser R^2 para que el modelo sea juzgado como útil al nivel de significancia .05?

81. ¿La exposición a la contaminación del aire provoca una menor esperanza de vida? Esta pregunta se examinó en el artículo "Does Air Pollution Shorten Lives?" (*Statistics and Public Policy*, Reading, MA. Addison-Wesley, 1977). Los datos sobre

y = porcentaje total de mortalidad (muertes por 10,000)

x_1 = lectura media de partículas suspendidas ($\mu g/m^3$)

x_2 = lectura más baja de sulfato ($[\mu g/m^3)] \times 10$)

x_3 = densidad de población (personas/milla^2)

x_4 = (porcentaje no de raza blanca) $\times$ 10

x_5 = (porcentaje de más de 65 años) $\times$ 10

para el año 1960 se registraron para $n = 117$ áreas estadísticas metropolitanas estándar seleccionadas al azar. La ecuación de regresión estimada fue

$$y = 19.607 + .041x_1 + .071x_2 + .001x_3 + .041x_4 + .687x_5$$

a. Para este modelo, $R^2 = .827$. Usando un nivel de significancia de .05, efectúe una prueba de utilidad del modelo.

b. La desviación estándar estimada de $\hat{\beta}_1$ fue de .016. Calcule e interprete un intervalo de confianza de 90% para β_1.

c. Dado que la desviación estándar estimada para $\hat{\beta}_4$ es .007, determine si el porcentaje que no sea de raza blanca es una variable importante en el modelo. Utilice un nivel de significancia de .01.

d. En 1960, los valores de x_1, x_2, x_3, x_4, y x_5 para Pittsburgh fueron de 166, 60, 788, 68 y 95, respectivamente. Utilice la ecuación de regresión dada para predecir la tasa de mortalidad de Pittsburgh. ¿Cómo se compara la predicción del lector con el valor real de 1960 de 103 muertes por 10,000?

82. Dado que $R^2 = .723$ para el modelo que contiene los predictores x_1, x_4, x_5 y x_8 y $R^2 = .689$ para el modelo con predictores x_1, x_3, x_5 y x_6, ¿qué se puede decir acerca de R^2 para el modelo que contiene los predictores

a. x_1, x_3, x_4, x_5, x_6 y x_8? Explique.

b. x_1 y x_4? Explique.

Bibliografía

Chatterjee, Samprit, y Ali Hadi, *Regression Analysis by Example* (4a. ed.), Wiley, Nueva York, 2006. Una breve pero informativa discusión de temas seleccionados, en especial multicolinealidad y el uso de métodos sesgados de estimación.

Daniel, Cuthbert, y Fred Wood, *Fitting Equations to Data* (2a. Ed.), Wiley. Nueva York, 1980. Contiene muchas ideas y métodos que evolucionaron de la gran experiencia de consulta de los autores.

Draper, Norman, y Harry Smith, *Applied Regression Analysis* (3a. ed.) Wiley. Nueva York, 1999. Vea la bibliografía del capítulo 12.

Hoaglin, David, y Roy Welsch, "The Hat Matrix in Regression and ANOVA", *American Statistician*, 1978: 17–23. Describe métodos para detectar observaciones influyentes en un conjunto de datos de regresión.

Hocking, Ron, "The Analysis and Selection of Variables in Linear Regression", *Biometrics*, 1976: 1–49. Un excelente examen de este tema.

Neter, John, Michael Kutner, Christopher Nachtsheim, y William Wasserman, *Applied Linear Statistical Models* (5a. ed.), Irwin, Homewood, IL, 2004. Vea la bibliografía del capítulo 12.

14 Pruebas de bondad de ajuste y análisis de datos categóricos

INTRODUCCIÓN

En el tipo más sencillo de situación considerado en este capítulo, cada observación en una muestra se clasifica como perteneciente a uno de un número finito de categorías (por ejemplo, el tipo de sangre podría ser una de cuatro categorías O, A, B o AB). Con p_i se denota la probabilidad de que cualquier observación particular pertenezca a la categoría i (o la proporción de la población que pertenece a la categoría i). Se desea probar una hipótesis nula que satisfaga por completo los valores de todas las p_i (por ejemplo H_0: $p_1 = .45$, $p_2 = .35$, $p_3 = .15$, $p_4 = .05$, cuando hay cuatro categorías). El estadístico de prueba será una medida de la discrepancia entre los números observados de las categorías y los correspondientes números esperados cuando H_0 es verdadera. Debido a que se llegará a una decisión al comparar el valor del estadístico de prueba para un valor crítico de la distribución chi cuadrada, el procedimiento recibe el nombre de prueba de bondad de ajuste chi cuadrada.

A veces la hipótesis nula especifica que las p_i dependen de algún número más pequeño de parámetros sin especificar los valores de estos parámetros. Por ejemplo, con tres categorías la hipótesis nula podría indicar que $p_1 = \theta^2$, $p_2 = 2\theta(1 - \theta)$, y $p_3 = (1 - \theta)^2$. Para efectuar una prueba de chi cuadrada, los valores de cualesquiera parámetros no especificados deben estimarse a partir de datos muestrales. Estos problemas se estudian en la sección 14.2. Los métodos se aplican entonces para probar una hipótesis nula que exprese que la muestra proviene de una familia particular de distribuciones, como la familia Poisson (con μ estimada desde la muestra) o la familia normal (con μ y σ estimadas). En resumen, una prueba basada en una gráfica de probabilidad normal es presentada para la hipótesis nula de la normalidad de la población.

Las pruebas de chi cuadrada para dos situaciones diferentes se presentan en la sección 14.3. En la primera, la hipótesis nula expresa que las p_i son iguales para varias poblaciones diferentes. El segundo tipo de situación comprende el tomar una muestra de una población individual y clasificar a cada individuo con respecto a dos factores categóricos diferentes (por ejemplo, la preferencia religiosa y el registro de partido político). La hipótesis nula en esta situación es que los dos factores son independientes dentro de la población.

14.1 Pruebas de bondad de ajuste cuando las probabilidades categóricas se satisfacen por completo

Un experimento binomial consiste en una secuencia de intentos independientes en la que cada intento puede resultar en uno de dos posibles resultados, S (por éxito) y F (por fracaso). La probabilidad de éxito, denotada por p, se supone constante de un intento a otro, y el número n de intentos se fija al principio del experimento. En el capítulo 8 se presentó una prueba z de muestra grande para probar H_0: $p = p_0$. Note que esta hipótesis nula especifica $P(S)$ y $P(F)$, porque si $P(S) = p_0$, entonces $P(F) = 1 - p_0$. Si se denota $P(F)$ por q y $1 - p_0$ por q_0, la hipótesis nula se puede escribir alternativamente como H_0: $p = p_0$, $q = q_0$. La prueba z es de dos colas cuando la alternativa de interés es $p \neq p_0$.

Un **experimento multinomial** generaliza un experimento binomial al permitir que cada intento resulte en uno de k posibles resultados, donde $k > 2$. Por ejemplo, suponga que una tienda acepta tres tipos diferentes de tarjetas de crédito. Un experimento multinomial resultaría de observar el tipo de tarjeta de crédito que utilizan, ya sea tipo 1, tipo 2 o tipo 3, cada uno de los siguientes n clientes que pagan con tarjeta de crédito. En general, nos referiremos a los k posibles resultados en una tirada dada como categorías y p_i denotará la probabilidad de un resultado en dicha categoría i. Si el experimento consiste en seleccionar n individuos u objetos de una población y clasificar cada uno, entonces p_i es la proporción de la población que cae en la i-ésima categoría (un experimento como éste será aproximadamente multinomial siempre que n sea mucho menor que el tamaño de la población).

La hipótesis nula de interés especificará el valor de cada p_i. Por ejemplo, en el caso de $k = 3$, se podría tener H_0: $p_1 = .5$, $p_2 = .3$, $p_3 = .2$. La hipótesis alternativa indicará que H_0 no es verdadera, es decir, que al menos una de las p_i tiene un valor diferente de lo expresado por H_0 (en cuyo caso al menos dos deben ser distintas, porque deben sumar 1). El símbolo p_{i0} representará el valor de p_i indicado por la hipótesis nula. En el ejemplo que acabamos de ver, $p_{10} = .5$, $p_{20} = .3$ y $p_{30} = .2$.

Antes de llevar a cabo el experimento multinomial, el número de intentos que da por resultado la categoría i ($i = 1, 2, \ldots,$ o k) es una variable aleatoria, en la misma forma que el número de éxitos y el número de fracasos en un experimento binomial son variables aleatorias. Esta variable aleatoria estará denotada por N_i y su valor observado por n_i. En vista que cada intento da por resultado exactamente una de las k categorías, $\sum N_i = n$ y lo mismo es verdadero de las n_i. Como ejemplo, un experimento con $n = 100$ y $k = 3$ podría dar $N_1 = 46$, $N_2 = 35$ y $N_3 = 19$.

El número esperado de éxitos y el número esperado de fracasos en un experimento binomial son np y nq, respectivamente. Cuando H_0: $p = p_0$, $q = q_0$ es verdadera, los números esperados de éxitos y fracasos son np_0 y nq_0, respectivamente. Del mismo modo, en un experimento multinomial el número esperado de intentos que resulte en la categoría i es $E(N_i) = np_i$ ($i = 1, \ldots, k$). Cuando H_0: $p_1 = p_{10}, \ldots, p_k = p_{k0}$ es verdadera, estos valores esperados se convierten en $E(N_1) = np_{10}$, $E(N_2) = np_{20}, \ldots, E(N_k) = np_{k0}$. Para el caso en que $k = 3$, H_0: $p_1 = .5$, $p_2 = .3$, $p_3 = .2$ y $n = 100$, las frecuencias esperadas cuando H_0 es verdadera son $E(N_1) = 100(.5) = 50$, $E(N_2) = 30$ y $E(N_3) = 20$. Las n_i y sus corres-

pondientes frecuencias esperadas a menudo son mostradas en formato tabular como se ilustra en la tabla 14.1. Los valores esperados cuando H_0 es verdadera se muestran justo debajo de los valores observados. Las N_i y las n_i suelen llamarse *cantidades de celda observadas* (o *frecuencias de celda observadas*), y $np_{10}, np_{20}, \ldots, np_{k0}$ son las correspondientes *cantidades de celda esperadas* bajo H_0.

Tabla 14.1 Cantidades de celda observadas y esperadas

Categoría	$i = 1$	$i = 2$	. . .	$i = k$	Total de fila
Observada	n_1	n_2	. . .	n_k	n
Esperada	np_{10}	np_{20}	. . .	np_{k0}	n

Las n_i deben ser razonablemente cercanas a las correspondientes np_{i0} cuando H_0 es verdadera. Por otra parte, varias de las cantidades observadas deben diferir en forma considerable de estas cantidades esperadas cuando los valores reales de las p_i difieran marcadamente de lo que indica la hipótesis nula. El procedimiento de prueba comprende evaluar la discrepancia entre las n_i y las np_{i0}, con H_0 rechazada cuando la discrepancia es lo suficientemente grande. Es natural basar una medida de discrepancia en los cuadrados de las desviaciones $(n_1 - np_{10})^2, (n_2 - np_{20})^2, \ldots, (n_k - np_{k0})^2$. Una forma aparente de combinar éstas en una medida general es sumarlas para obtener $\sum(n_i - np_{i0})^2$. No obstante, suponga que $np_{10} = 100$ y $np_{20} = 10$. Entonces si $n_1 = 95$ y $n_2 = 5$, las dos categorías contribuyen con las mismas desviaciones al cuadrado a la medida propuesta. Pero n_1 es sólo 5% menor de lo que se esperaría cuando H_0 es verdadera, mientras que n_2 es 50% menor. Para tomar en cuenta las magnitudes relativas de las desviaciones, cada desviación cuadrática se divide entre la correspondiente cantidad esperada.

Antes de dar una descripción más detallada, se debe analizar un tipo de distribución de probabilidad llamada *distribución chi cuadrada*. Esta distribución se introdujo primero en la sección 4.4 y se usó en el capítulo 7 para obtener un intervalo de confianza para la varianza σ^2 de una población normal. La distribución chi cuadrada tiene un solo parámetro ν, llamado número de grados de libertad (gl) de la distribución, con posibles valores 1, 2, 3, . . . Análogo al valor crítico $t_{\alpha,\nu}$ para la distribución t, $\chi^2_{\alpha,\nu}$ es el valor tal que α del área bajo la curva χ^2 con grado de libertad ν está a la derecha de $\chi^2_{\alpha,\nu}$ (vea la figura 14.1). Valores seleccionados de $\chi^2_{\alpha,\nu}$ se dan en la tabla A.7 del apéndice.

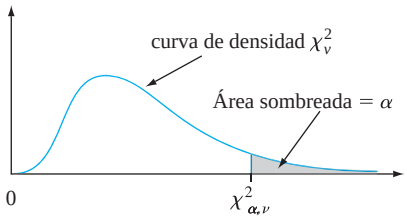


Figura 14.1 Un valor crítico para una distribución chi cuadrada

TEOREMA

Siempre que $np_i \geq 5$ para toda i $(i = 1, 2, \ldots, k)$, la variable

$$\chi^2 = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i} = \sum_{\text{todas las celdas}} \frac{(\text{observada} - \text{esperada})^2}{\text{esperada}}$$

tiene aproximadamente una distribución chi cuadrada con $k - 1$ grados de libertad.

El hecho de que gl $= k - 1$ es una consecuencia de la restricción $\sum N_i = n$. Aun cuando haya k cuentas de celdas observadas, una vez conocida cualquier $k - 1$, la restante se determina de manera única. Esto es, hay sólo $k - 1$ cuentas de celdas "determinadas libremente" y por tanto $k - 1$ grados de libertad.

Si np_{i0} se sustituye por np_i en χ^2, la estadística de prueba resultante tiene una distribución chi cuadrada cuando H_0 es verdadera. El rechazo de H_0 es apropiado cuando $\chi^2 \geq c$ (porque grandes discrepancias entre cuentas observadas y esperadas llevan a un valor grande de χ^2), y la opción $c = \chi^2_{\alpha,k-1}$ da una prueba con nivel de significancia α.

Hipótesis nula: H_0: $p_1 = p_{10}, p_2 = p_{20}, \ldots, p_k = p_{k0}$

Hipótesis alternativa: H_a: al menos una p_i no es igual a p_{i0}

Valor de estadística de prueba: $\chi^2 = \sum_{\substack{\text{todas} \\ \text{las celdas}}} \dfrac{(\text{observado} - \text{esperado})^2}{\text{esperado}} = \sum_{i=1}^{k} \dfrac{(n_i - np_{i0})^2}{np_{i0}}$

Región de rechazo: $\chi^2 \geq \chi^2_{\alpha,k-1}$

Ejemplo 14.1 Si se concentra en dos características diferentes de un organismo, cada uno controlado por un solo gen, y se cruza una variedad pura que tenga genotipo *AABB* con una variedad pura que tenga genotipo *aabb* (las letras mayúsculas denotan alelos dominantes; las minúsculas, alelos recesivos), el genotipo resultante será *AaBb*. Si estos organismos de primera generación se cruzan entonces entre ellos (un cruce dihíbrido), habrá cuatro fenotipos que dependen de si está presente un alelo dominante de uno de los dos tipos. Las leyes de Mendel de la herencia implican que estos cuatro fenotipos deben tener probabilidades $\frac{9}{16}, \frac{3}{16}, \frac{3}{16}$ y $\frac{1}{16}$ de aparecer en cualquier cruce dihíbrido determinado.

El artículo "Linkage Studies of the Tomato" (*Trans. Royal Canadian Institute*, 1931: 1–19) publica los siguientes datos sobre fenotipos de un cruce dihíbrido de tomates altos de hoja cortada con tomates enanos de hoja de papa. Hay $k = 4$ categorías que corresponden a los cuatro fenotipos posibles, con la hipótesis nula

$$H_0: p_1 = \frac{9}{16},\ p_2 = \frac{3}{16},\ p_3 = \frac{3}{16},\ p_4 = \frac{1}{16}$$

Las cantidades esperadas de celdas son $9n/16$, $3n/16$, $3n/16$, y $n/16$ y la prueba está basada en $k - 1 = 3$ grados de libertad. El tamaño muestral total fue $n = 1611$. Las cantidades observadas y esperadas se dan en la tabla 14.2.

Tabla 14.2 Cantidades observadas y esperadas de celdas para el ejemplo 14.1

	$i = 1$ Alto, hoja cortada	$i = 2$ Alto, hoja de papa	$i = 3$ Enano, hoja cortada	$i = 4$ Enano, hoja de papa
n_i	926	288	293	104
np_{i0}	906.2	302.1	302.1	100.7

La contribución a χ^2 desde la primera celda es

$$\frac{(n_1 - np_{10})^2}{np_{10}} = \frac{(926 - 906.2)^2}{906.2} = .433$$

Las celdas 2, 3, y 4 contribuyen con .658, .274 y .108, respectivamente, de modo que $\chi^2 = .433 + .658 + .274 + .108 = 1.473$. Una prueba con nivel de significancia .10 requiere $\chi^2_{.10,\,3}$, el número de la fila de grado de libertad 3 y la columna .10 de la tabla A.7 del apéndice. Este valor crítico es 6.251. Como 1.473 no es al menos 6.251, H_0 no puede ser rechazada incluso a este nivel más bien grande de significación. La información es bastante consistente con las leyes de Mendel. ■

Aún cuando hemos desarrollado la prueba de chi cuadrada para situaciones en las que $k > 2$, también se puede usar cuando $k = 2$. La hipótesis nula en este caso se puede expresar como H_0: $p_1 = p_{10}$, porque las relaciones $p_2 = 1 - p_1$ y $p_{20} = 1 - p_{10}$ hacen redundante la inclusión de $p_2 = p_{20}$ en H_0. La hipótesis alternativa es H_a: $p_1 \neq p_{10}$. Estas hipótesis también se pueden probar usando una prueba z de dos colas con estadístico de prueba

$$Z = \frac{(N_1/n) - p_{10}}{\sqrt{\dfrac{p_{10}(1 - p_{10})}{n}}} = \frac{\hat{p}_1 - p_{10}}{\sqrt{\dfrac{p_{10}p_{20}}{n}}}$$

Sorprendentemente, los dos procedimientos son por completo equivalentes. Esto es porque se puede demostrar que $Z^2 = \chi^2$ y $(z_{\alpha/2})^2 = \chi^2_{1,\alpha}$, de modo que $\chi^2 \geq \chi^2_{1,\alpha}$ si y sólo si $|Z| \geq z_{\alpha/2}$.* Si la hipótesis alternativa es H_a: $p_1 > p_{10}$ o H_a: $p_1 < p_{10}$, la prueba chi cuadrada no se puede usar. Se debe entonces revertir a una prueba z de cola superior o inferior.

Al igual que en el caso con todos los procedimientos de prueba, se debe tener cuidado de no confundir significación estadística con significación práctica. Una χ^2 calculada que exceda de $\chi^2_{\alpha,k-1}$ puede resultar de un tamaño muestral muy grande más que por cualesquiera diferencias prácticas entre las p_{i0} hipotéticas y las verdaderas p_i. Entonces, si $p_{10} = p_{20} = p_{30} = \frac{1}{3}$, pero las p_i verdaderas tienen valores de .330, .340 y .330, es seguro que aparecerá un valor grande de χ^2 con una n que sea suficientemente grande. Antes de rechazar H_0, las $\hat{p}_i$ deben examinarse para ver si sugieren un modelo diferente del de H_0 desde un punto de vista práctico.

Valores *P* para pruebas chi cuadrada

Las pruebas chi cuadrada en este capítulo son todas de cola superior, de modo que ése será el enfoque. Así como el valor P para una prueba t de cola superior es el área bajo la curva t_ν a la derecha de la t calculada, el valor P para una prueba chi cuadrada de cola superior es el área bajo la curva χ^2_ν a la derecha de la χ^2 calculada. La tabla A.7 del apéndice da información limitada del valor P porque sólo se tabulan cinco valores críticos de cola superior para cada ν diferente. Por tanto, se incluye otra tabla de apéndice, análoga a la tabla A.8, que facilita hacer enunciados más precisos de valor P.

El hecho de que las curvas t estuvieran todas centradas en cero permitió tabular áreas de cola de curva t en forma relativamente compacta, con el margen izquierdo dando valores que van de 0.0 a 4.0 en la escala t horizontal y varias columnas que muestran áreas correspondientes de cola superior para varios grados de libertad. El movimiento hacia la derecha de curvas de chi cuadrada cuando aumenta el grado de libertad necesita un tipo de tabulación un poco diferente. El margen izquierdo de la tabla A.11 del apéndice da varias áreas de cola superior: .100, .095, .090, . . . , .005 y .001. Cada columna de la tabla es para un valor

* El hecho que $(z_{\alpha/2})^2 = \chi^2_{1,\alpha}$ es una consecuencia de la relación entre la distribución normal estándar y la distribución chi cuadrada con grado de libertad 1; si $Z \sim N(0,1)$, entonces Z^2 tiene una distribución chi cuadrada con $\nu = 1$.

diferente de grado de libertad, y las entradas son valores sobre el eje horizontal chi cuadrada que capta estas áreas correspondientes de cola. Por ejemplo, al moverse hacia abajo al área de cola .085, y en sentido horizontal a la columna de grado de libertad 4, se ve que el área a la derecha de 8.18 bajo la curva chi cuadrada de grado de libertad 4 es .085 (vea la figura 14.2).

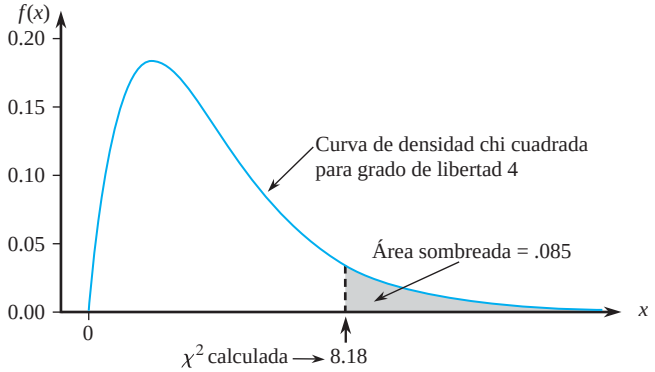


Figura 14.2 Un valor P para una prueba chi cuadrada de cola superior

Para captar esta misma área de cola superior bajo la curva de grado de libertad 10, se debe salir a 16.54. En la columna de grado de libertad 4, la fila superior muestra que si el valor calculado de la variable chi cuadrada es menor a 7.77, el área de cola captada (el valor P) excede de .10. Del mismo modo, la fila inferior en esta columna indica que si el valor calculado excede de 18.46, el área de cola es menor a .001 (el valor $P < .001$).

χ^2 cuando las p_i son funciones de otros parámetros

Es frecuente que las p_i sean hipotéticas para depender de un número menor de parámetros $\theta_1, \ldots, \theta_m$ $(m < k)$. Entonces una hipótesis específica que comprenda las θ_i da las p_{i0} específicas, que entonces se usan en la prueba χ^2.

Ejemplo 14.2 En un bien conocido artículo sobre genética ("The Progeny in Generations F_{12} to F_{17} of a Cross Between a Yellow-Wrinkled and a Green-Round Seeded Pea", *J. of Genetics*, 1923: 255–331), G. U. Yule, uno de los primeros expertos en estadística, analizó datos que resultaban de cruzar chícharos producidos en un jardín. Los alelos dominantes en el experimento fueron Y = color amarillo y R = forma redonda, que dieron por resultado el YR dominante doble. Yule examinó 269 vainas de cuatro semillas que resultaron de un cruce dihíbrido y contó el número de semillas YR de cada vaina. Denotando con X el número de los YR de una vaina seleccionada al azar, los posibles valores X son 0, 1, 2, 3, 4, que se identifican con las celdas 1, 2, 3, 4 y 5 de una tabla rectangular (entonces, por ejemplo, una vaina con $X = 4$ da una cantidad observada en la celda 5).

La hipótesis de que las leyes de Mendel son operativas, y que los genotipos de semillas individuales dentro de una vaina son independientes entre sí, implica que X tiene una distribución binomial con $n = 4$ y $\theta = \frac{9}{16}$. Entonces se desea probar H_0: $p_1 = p_{10}, \ldots, p_5 = p_{50}$, donde

$$p_{i0} = P(i - 1 \text{ YR entre 4 semillas cuando } H_0 \text{ es verdadera})$$
$$= \binom{4}{i-1} \theta^{i-1}(1-\theta)^{4-(i-1)} \qquad i = 1, 2, 3, 4, 5;\ \theta = \frac{9}{16}$$

Los datos y cálculos de Yule están en la tabla 14.3 con cantidades de celda esperadas $np_{i0} = 269p_{i0}$.

Tabla 14.3 Cantidades de celdas observadas y esperadas para el ejemplo 14.2

Celda *i*	**1**	**2**	**3**	**4**	**5**
Chícharos YR por vaina	**0**	**1**	**2**	**3**	**4**
Observadas	16	45	100	82	26
Esperadas	9.86	50.68	97.75	83.78	26.93
$\frac{(\text{observadas} - \text{esperadas})^2}{\text{esperadas}}$	3.823	.637	.052	.038	.032

Entonces $\chi^2 = 3.823 + \cdots + .032 = 4.582$. Debido a que $\chi^2_{.01,k-1} = \chi^2_{.01,4} = 13.277$, H_0 no es rechazada al nivel .01. La tabla A.11 del apéndice muestra que como $4.582 < 7.77$, el valor P para la prueba excede de .10. H_0 no debe ser rechazada a ningún nivel razonable de significación. ■

χ^2 cuando la distribución básica es continua

Hasta aquí se ha supuesto que las k categorías están naturalmente definidas en el contexto del experimento bajo consideración. La prueba χ^2 también se puede usar para probar si una muestra proviene de una distribución continua básica específica. Denótese con X la variable que se muestrea y supóngase que la función de densidad de probabilidad (fdp) hipotética de X es $f_0(x)$. Al igual que en la construcción de una distribución de frecuencia en el capítulo 1, subdivida la escala de medición de X en k intervalos $[a_0, a_1), [a_1, a_2), \ldots, [a_{k-1}, a_k)$, donde el intervalo $[a_{i-1}, a_i)$ incluye el valor a_{i-1} pero no a_i. Las probabilidades de celda especificadas por H_0 son entonces

$$p_{i0} = P(a_{i-1} \le X < a_i) = \int_{a_{i-1}}^{a_i} f_0(x)\,dx$$

Las celdas deben escogerse de modo que $np_{i0} \ge 5$ para $i = 1, \ldots, k$. Es frecuente que se seleccionen de modo que las np_{i0} sean iguales.

Ejemplo 14.3 Para ver si el tiempo de inicio de trabajo de parto en madres está uniformemente distribuido en todo un día de 24 horas, se puede dividir un día en k periodos, cada uno con duración $24/k$. La hipótesis nula expresa que $f(x)$ es la función de densidad de probabilidad uniforme en el intervalo [0, 24], de modo que $p_{i0} = 1/k$. El artículo "The Hour of Birth" (*British J. of Preventive and Social Medicine*, 1953: 43–59) habla de 1186 tiempos de inicio, que se clasificaron en $k = 24$ intervalos de 1 hora que principiaban a la media noche, y resultaron en cuentas de celda de 52, 73, 89, 88, 68, 47, 58, 47, 48, 53, 47, 34, 21, 31, 40, 24, 37, 31, 47, 34, 36, 44, 78 y 59. Cada una de las cantidades esperadas de celda es $1186 \cdot \frac{1}{24} = 49.42$ y el valor resultante de χ^2 es 162.77. Como $\chi^2_{.01,23} = 41.637$, el valor calculado es muy significativo, y la hipótesis nula tiene un rotundo rechazo. Hablando en términos generales, parece que es mucho más probable que el trabajo de parto se inicie ya bien entrada la noche que durante las horas hábiles normales. ■

Para probar si una muestra proviene de una distribución normal específica, los parámetros fundamentales son $\theta_1 = \mu$ y $\theta_2 = \sigma$, y cada p_{i0} será una función de estos parámetros.

Ejemplo 14.4 En cierta universidad, se supone que los exámenes finales duran 2 horas. El departamento de psicología construyó un examen final departamental para un curso elemental que se pensaba podría satisfacer los siguientes criterios: (1) el tiempo real tomado para completar el examen está distribuido de manera normal, (2) $\mu = 100$ min y (3) exactamente 90%

de todos los estudiantes terminarán dentro del periodo de 2 horas. Para ver si éste es el caso en realidad, se seleccionaron al azar 120 estudiantes y se registraron sus tiempos para terminar el examen. Se decidió que debían emplearse $k = 8$ intervalos. Los criterios implican que el 90º percentil de la distribución del tiempo de terminación es $\mu + 1.28\sigma = 120$. Como $\mu = 100$, esto implica que $\sigma = 15.63$.

Los ocho intervalos que dividen la escala normal estándar en ocho segmentos igualmente probables son [0, .32), [.32, .675), [.675, 1.15) y [1.15, ∞), y sus cuatro similares en el otro lado de 0. Para $\mu = 100$ y $\sigma = 15.63$, estos intervalos se convierten en [100, 105), [105, 110.55), [110.55, 117.97) y [117.97, ∞). Así, $p_{i0} = \frac{1}{8} = .125$ $(i = 1, \ldots, 8)$, de modo que cada cantidad de celdas esperada es $np_{i0} = 120(.125) = 15$. Las cantidades de celda observadas fueron 21, 17, 12, 16, 10, 15, 19 y 10, que dan por resultado una χ^2 de 7.73. Como $\chi^2_{.10,7} = 12.017$ y 7.73 no es ≥ 12.017, no hay evidencia para concluir que los criterios no se han satisfecho. ■

EJERCICIOS Sección 14.1 (1–11)

1. ¿Cuál conclusión sería apropiada para una prueba de chi cuadrada de cola superior en cada una de las situaciones siguientes?

a. $\alpha = .05$, gl $= 4$, $\chi^2 = 12.25$
b. $\alpha = .01$, gl $= 3$, $\chi^2 = 8.54$
c. $\alpha = .10$, gl $= 2$, $\chi^2 = 4.36$
d. $\alpha = .01$, $k = 6$, $\chi^2 = 10.20$

2. Diga cuanto pueda acerca del valor *P* para una prueba de chi cuadrada de cola superior en cada una de las situaciones siguientes.

a. $\chi^2 = 7.5$, gl $= 2$ **b.** $\chi^2 = 13.0$, gl $= 6$
c. $\chi^2 = 18.0$, gl $= 9$ **d.** $\chi^2 = 21.3$, gl $= 5$
e. $\chi^2 = 5.0$, $k = 4$

3. El artículo "Racial Stereotypes in Children's Television Commercials" (*J. of Adver. Res.*, 2008: 80–93) reportaron las siguientes frecuencias con las que los caracteres étnicos aparecieron en anuncios registrados que salieron al aire en las estaciones de televisión de Filadelfia.

Grupo étnico:	Afro-americanos	Asiáticos	Caucásicos	Hispanos
Frecuencia	57	11	330	6

En el censo del año 2000, las proporciones de estos cuatro grupos étnicos son .177, .032, .734 y .057, respectivamente. ¿Los datos sugieren que la proporción de anuncios es diferente de las proporciones del censo? Lleve a cabo una prueba de hipótesis apropiada con un nivel de significancia de .01 y también diga lo más que pueda sobre el valor *P*.

4. Se ha supuesto que cuando los pichones que regresan a su palomar se desorientan de algún modo, no mostrarán preferencia para ninguna dirección después de alzar el vuelo (de modo que la dirección *X* debería estar uniformemente distribuida en el intervalo de 0° a 360°). Para probar esto, se desorienta a 120 pichones, se sueltan y se registra la dirección de vuelo de cada uno de ellos; a continuación aparecen los datos resultantes. Utilice una prueba chi cuadrada al nivel .10 para ver si la información apoya a la hipótesis.

Dirección	0−<45°	45−<90°	90−<135°
Frecuencia	12	16	17
Dirección	135−< 180°	180−<225°	225−<270°
Frecuencia	15	13	20
Dirección	270−<315°	315−<360°	
Frecuencia	17	10	

5. Un sistema de recuperación de información tiene 10 lugares de almacenamiento. Se ha guardado información con la esperanza de que la proporción de peticiones a largo plazo para la posición *i*, sea dada por $p_i = (5.5 - |i - 5.5|)/30$. Una muestra de 200 peticiones de recuperación dio las siguientes frecuencias para las posiciones 1–10, respectivamente: 4, 15, 23, 25, 38, 31, 32, 14, 10 y 8. Utilice una prueba de chi cuadrada al nivel de significancia de .10 para determinar si la información es consistente con las proporciones *a priori* (use el método del valor *P*).

6. El artículo "The Gap Between Wine Expert Ratings and Consumer Preferences" (*Intl. J. of Wine Business Res.*, 2008: 335–351) estudió las diferencias entre las calificaciones de expertos y de los consumidores al considerar la medalla de evaluación para los vinos, la que podría ser de oro (G), plata (S) o de bronce (B). Tres categorías se establecieron a continuación: 1. La clasificación es la misma [(G, G), (B, B), (S, S)], 2. La clasificación difiere en una medalla [(G, S), (S, G), (S, B), (B, S)], y 3. La clasificación difiere por dos medallas [(G, B), (B, G)]. Las frecuencias observadas para estas tres categorías fueron 69, 102 y 45, respectivamente. En la hipótesis de las calificaciones igualmente probables de expertos y clasificaciones de los consumidores que se asignaron completamente al azar, cada uno de los nueve pares de medallas tiene una probabilidad de 1/9. Lleve a cabo una prueba adecuada de chi-cuadrada con un nivel de significancia de .10, obteniendo en primer lugar la información del valor *P*.

7. Durante mucho tiempo, los criminólogos han debatido sobre si hay una relación entre las condiciones climáticas y la incidencia

de delitos violentos. El autor del artículo "Is There a Season for Homicide?" (*Criminology*, 1988: 287–296) clasificó 1361 homicidios según la estación del año y resultaron los datos siguientes. Pruebe la hipótesis nula de iguales proporciones usando $\alpha = .01$ mediante el uso de la tabla chi cuadrada para decir cuanto sea posible acerca del valor P.

Invierno	Primavera	Verano	Otoño
328	334	372	327

8. El artículo "Psychiatric and Alcoholic Admissions Do Not Occur Disproportionately Close to Patients' Birthdays" (*Psychological Reports*, 1991: 944–946) se concentra en la existencia de alguna relación entre la fecha de ingreso de un paciente para tratamiento de alcoholismo y el cumpleaños del paciente. Suponiendo un año de 365 días (es decir, excluyendo un año bisiesto), en ausencia de cualquier relación, es probable que la fecha de admisión de un paciente sea cualquiera de los 365 días. Los investigadores establecieron cuatro diferentes categorías de admisión: (1) no más de 7 días de la fecha de cumpleaños, (2) entre 8 y 30 días, inclusive, desde el cumpleaños, (3) entre 31 y 90 días, inclusive, del cumpleaños, y (4) más de 90 días del cumpleaños. Una muestra de 200 pacientes dio frecuencias observadas de 11, 24, 69 y 96 para las categorías 1, 2, 3 y 4, respectivamente. Exprese y pruebe las hipótesis relevantes usando un nivel de significancia de .01.

9. Se ha teorizado el tiempo de respuesta de un sistema computarizado, a una petición de cierto tipo de información, como una distribución exponencial con parámetro $\lambda = 1$ segundo (de modo que si $X =$ tiempo de respuesta, la función de densidad de probabilidad de X bajo H_0 es $f_0(x) = e^{-x}$ para $x \geq 0$).

a. Si se hubiera observado $X_1, X_2, \ldots, X_n$ y se deseara usar la prueba de chi cuadrada con cinco intervalos de clase de igual probabilidad bajo H_0, ¿cuáles serían los intervalos de clase resultantes?

b. Realice la prueba de chi cuadrada usando los datos siguientes de una muestra aleatoria de 40 tiempos de respuesta:

.10	.99	1.14	1.26	3.24	.12	.26	.80
.79	1.16	1.76	.41	.59	.27	2.22	.66
.71	2.21	.68	.43	.11	.46	.69	.38
.91	.55	.81	2.51	2.77	.16	1.11	.02
2.13	.19	1.21	1.13	2.93	2.14	.34	.44

10. a. Demuestre que otra expresión para el estadístico de chi cuadrada es

$$\chi^2 = \sum_{i=1}^{k} \frac{N_i^2}{np_{i0}} - n$$

¿Por qué es más eficiente calcular χ^2 usando esta fórmula?

b. Cuando la hipótesis nula es H_0: $p_1 = p_2 = \cdots = p_k = 1/k$ (es decir, $p_{i0} = 1/k$ para toda i), ¿cómo se simplifica la fórmula del inciso (a)? Utilice la expresión simplificada para calcular χ^2 para los datos de pichón/dirección del ejercicio 4.

11. a. Habiendo obtenido una muestra aleatoria de una población, el lector desea usar una prueba chi cuadrada para determinar si la distribución de población es normal estándar. Si basa la prueba en seis intervalos de clase que tengan igual probabilidad bajo H_0, ¿cuáles deberían ser los intervalos de clase?

b. Si desea usar una prueba chi cuadrada para probar H_0: la distribución poblacional es normal con $\mu = .5$, $\sigma = .002$ y la prueba ha de basarse en seis intervalos de clase igualmente probables (bajo H_0). ¿Cuáles deben ser estos intervalos?

c. Use la prueba chi cuadrada con los intervalos del inciso (b) para determinar, con base en los siguientes 45 diámetros de tornillos, si el diámetro de éstos es una variable normalmente distribuida con $\mu = .5$ pulgada, $\sigma = .002$ pulgada.

.4974	.4976	.4991	.5014	.5008	.4993
.4994	.5010	.4997	.4993	.5013	.5000
.5017	.4984	.4967	.5028	.4975	.5013
.4972	.5047	.5069	.4977	.4961	.4987
.4990	.4974	.5008	.5000	.4967	.4977
.4992	.5007	.4975	.4998	.5000	.5008
.5021	.4959	.5015	.5012	.5056	.4991
.5006	.4987	.4968			

14.2 Pruebas de bondad de ajuste para hipótesis compuestas

En la sección previa, se presentó una prueba de bondad de ajuste basada en un estadístico χ^2 para decidir entre H_0: $p_1 = p_{10}, \ldots, p_k = p_{k0}$ y la alternativa H_a que expresa que H_0 no es verdadera. La hipótesis nula fue una **hipótesis simple** en el sentido que cada p_{i0} era un número especificado, de modo que las cantidades esperadas de celda cuando H_0 era verdadera eran números determinados de manera única.

En diversas situaciones, hay k categorías que ocurren naturalmente, pero H_0 expresa sólo que las p_i son funciones de otros parámetros $\theta_1, \ldots, \theta_m$ sin especificar los valores de estos parámetros θ. Por ejemplo, una población puede estar en equilibrio con respecto a las proporciones de los tres genotipos *AA*, *Aa* y *aa*. Con p_1, p_2, y p_3 denotando estas proporciones (probabilidades), se puede desear probar

$$H_0: p_1 = \theta^2,\ p_2 = 2\theta(1-\theta),\ p_3 = (1-\theta)^2 \qquad (14.1)$$

donde θ representa la proporción del gen A en la población. Esta hipótesis es **compuesta** porque saber que H_0 es verdadera no determina de manera única las probabilidades de

celda, y las cantidades esperadas de celda, sino sólo su forma general. Para llevar a cabo una prueba χ^2, los parámetros θ_i desconocidos deben estimarse en primer término.

Del mismo modo, puede haber interés en probar si una muestra provino de una familia particular de distribuciones sin especificar ningún miembro particular de la familia. Para usar la prueba χ^2 con el fin de ver si la distribución es Poisson, por ejemplo, debe estimarse el parámetro μ. Además, debido a que hay en realidad un número infinito de posibles valores de una variable Poisson, estos valores deben agruparse de manera que haya un número finito de celdas. Si H_0 expresa que la distribución básica es normal, el uso de una prueba χ^2 debe estar precedido por una selección de celdas y estimación de μ y σ.

χ^2 cuando se estiman parámetros

Al igual que antes, k denotará el número de categorías o celdas y p_i denotará la probabilidad de una observación que caiga en la i-ésima celda. La hipótesis nula indica ahora que cada p_i es una función de un pequeño número de parámetros $\theta_1, \ldots, \theta_m$ con los θ_i especificados de otro modo:

$$H_0\text{: } p_1 = \pi_1(\boldsymbol{\theta}), \ldots, p_k = \pi_k(\boldsymbol{\theta}) \qquad \text{donde } \boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$$

$$H_a\text{: la hipótesis } H_0 \text{ no es verdadera} \tag{14.2}$$

Por ejemplo, para H_0 de (14.1), $m = 1$ (hay sólo un θ), $\pi_1(\theta) = \theta^2$, $\pi_2(\theta) = 2\theta(1 - \theta)$ y $\pi_3(\theta) = (1 - \theta)^2$.

En el caso de $k = 2$, realmente hay sólo una variable aleatoria, N_1 (porque $N_1 + N_2 = n$), que tiene una distribución binomial. La probabilidad conjunta de que $N_1 = n_1$ y $N_2 = n_2$ es entonces

$$P(N_1 = n_1, N_2 = n_2) = \binom{n}{n_1} p_1^{n_1} \cdot p_2^{n_2} \propto p_1^{n_1} \cdot p_2^{n_2}$$

donde $p_1 + p_2 = 1$ y $n_1 + n_2 = n$. Para k general, la distribución conjunta de $N_1, \ldots, N_k$ es la distribución multinomial (sección 5.1) con

$$P(N_1 = n_1, \ldots, N_k = n_k) \propto p_1^{n_1} \cdot p_2^{n_2} \cdot \cdots \cdot p_k^{n_k} \tag{14.3}$$

Cuando H_0 es verdadera, (14.3) se convierte en

$$P(N_1 = n_1, \ldots, N_k = n_k) \propto [\pi_1(\theta)]^{n_1} \cdot \cdots \cdot [\pi_k(\theta)]^{n_k} \tag{14.4}$$

Para aplicar una prueba chi cuadrada, debe estimarse $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$.

MÉTODO DE ESTIMACIÓN

Denótese con $n_1, n_2, \ldots, n_k$ los valores observados de $N_1, \ldots, N_k$. Entonces $\hat{\theta}_1, \ldots, \hat{\theta}_m$ son los valores de los θ_i que maximizan (14.4).

Los estimadores resultantes $\hat{\theta}_1, \ldots, \hat{\theta}_m$ son los estimadores de máxima probabilidad de $\theta_1, \ldots, \theta_m$; este principio de estimación se discutió en la sección 6.2.

Ejemplo 14.5 En seres humanos hay un grupo sanguíneo, el MN, compuesto por personas que tienen uno de los tres tipos de sangre M, MN y N. El tipo está determinado por dos alelos y no hay predominio, de modo que los tres posibles genotipos dan lugar a tres fenotipos. Una población formada por personas del grupo MN está en equilibrio si

$$P(\text{M}) = p_1 = \theta^2$$

$$P(\text{MN}) = p_2 = 2\theta(1 - \theta)$$

$$P(\text{N}) = p_3 = (1 - \theta)^2$$

para algún θ. Suponga que una muestra de esa población dio los resultados mostrados en la tabla 14.4.

Tabla 14.4 Cantidades observadas para el ejemplo 14.5

Tipo	M	MN	N	
Observado	125	225	150	$n = 500$

Entonces

$$[\pi_1(\theta)]^{n_1}[\pi_2(\theta)]^{n_2}[\pi_3(\theta)]^{n_3} = [(\theta^2)]^{n_1}[2\theta(1-\theta)]^{n_2}[(1-\theta)^2]^{n_3}$$
$$= 2^{n_2} \cdot \theta^{2n_1+n_2} \cdot (1-\theta)^{n_2+2n_3}$$

Al maximizar esto con respecto a θ (o bien, lo que es lo mismo, al maximizar el logaritmo natural de esta cantidad, que es más fácil de derivar) se obtiene

$$\hat{\theta} = \frac{2n_1 + n_2}{[(2n_1 + n_2) + (n_2 + 2n_3)]} = \frac{2n_1 + n_2}{2n}$$

Con $n_1 = 125$ y $n_2 = 225$, $\hat{\theta} = 475/1000 = .475$. ■

Una vez que $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$ haya sido estimado por $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_m)$, las cantidades de células esperadas estimadas son los $n\pi_i(\hat{\theta})$. Éstos se usan ahora en lugar de los np_{i0} de la sección 14.1 para especificar un estadístico χ^2.

TEOREMA

Bajo condiciones generales de "regularidad" en $\theta_1, \ldots, \theta_m$ y las $\pi_i(\boldsymbol{\theta})$, si $\theta_1, \ldots, \theta_m$ se estiman por el método de máxima probabilidad como se describió antes y n es grande,

$$\chi^2 = \sum_{\text{todas las celdas}} \frac{(\text{observada} - \text{esperada estimada})^2}{\text{esperada estimada}} = \sum_{i=1}^{k} \frac{[N_i - n\pi_i(\hat{\boldsymbol{\theta}})]^2}{n\pi_i(\hat{\boldsymbol{\theta}})}$$

tiene aproximadamente una distribución chi cuadrada con $k - 1 - m$ grados de libertad cuando H_0 de (14.2) es verdadera. Una prueba α aproximadamente nivelada de H_0 contra H_a es entonces para rechazar H_0 si $\chi^2 \geq \chi^2_{\alpha, k-1-m}$. En la práctica, la prueba se puede usar si $n\pi_i(\hat{\boldsymbol{\theta}}) \geq 5$ para toda i.

Note que *el número de grados de libertad es reducido por el número de los θ_i estimados.*

Ejemplo 14.6 (Continuación del ejemplo 14.5)

Con $\hat{\theta} = .475$ y $n = 500$, las cantidades de celda esperadas estimadas son $n\pi_1(\hat{\theta}) = 500(\hat{\theta})^2 = 112.81$, $n\pi_2(\hat{\theta}) = (500)(2)(.475)(1-.475) = 249.38$ y $n\pi_3(\hat{\theta}) = 500 - 112.81 - 249.38 = 137.81$. Entonces

$$\chi^2 = \frac{(125 - 112.81)^2}{112.81} + \frac{(225 - 249.38)^2}{249.38} + \frac{(150 - 137.81)^2}{137.81} = 4.78$$

Como $\chi^2_{.05,\ k-1-m} = \chi^2_{.05,\ 3-1-1} = \chi^2_{.05,\ 1} = 3.843$ y $4.78 \geq 3.843$, H_0 es rechazada. La tabla A.11 del apéndice muestra que el valor $P \approx .029$. ■

Ejemplo 14.7

Considere una serie de juegos entre dos equipos, I y II, que termina tan pronto como un equipo haya ganado cuatro juegos (sin posibilidad de empate). Un modelo simple de probabilidad para esa serie supone que los resultados de juegos sucesivos son independientes,

y que la probabilidad de que el equipo I gane cualquier juego en particular es una constante θ. De manera arbitraria se designa al equipo I como el mejor, de modo que $\theta \geq .5$. Cualquier serie particular puede terminar entonces después de 4, 5, 6 o 7 juegos. Sean $\pi_1(\theta)$, $\pi_2(\theta)$, $\pi_3(\theta)$, $\pi_4(\theta)$ que denotan la probabilidad de terminar en 4, 5, 6 y 7 juegos, respectivamente. Entonces

$$\begin{aligned}
\pi_1(\theta) &= P(\text{I gana en 4 juegos}) + P(\text{II gana en 4 juegos}) \\
&= \theta^4 + (1-\theta)^4 \\
\pi_2(\theta) &= P(\text{I gana 3 de los primeros 4 y el quinto}) \\
&\quad + P(\text{I pierde 3 de los primeros 4 y el quinto}) \\
&= \binom{4}{3}\theta^3(1-\theta)\cdot\theta + \binom{4}{1}\theta(1-\theta)^3\cdot(1-\theta) \\
&= 4\theta(1-\theta)[\theta^3 + (1-\theta)^3] \\
\pi_3(\theta) &= 10\theta^2(1-\theta)^2[\theta^2 + (1-\theta)^2] \\
\pi_4(\theta) &= 20\theta^3(1-\theta)^3
\end{aligned}$$

El artículo "Seven-Game Series in Sports" de Groeneveld y Meeden (*Mathematics Magazine*, 1975: 187-192) probó el ajuste de este modelo a los resultados de juegos para determinar el campeonato de la National Hockey League, durante el periodo 1943–1967 (cuando la membresía de la liga era estable). Los datos aparecen en la tabla 14.5.

Tabla 14.5 Cantidades observadas y esperadas para el modelo simple

Celda	**1**	**2**	**3**	**4**	
Número de juegos jugados	**4**	**5**	**6**	**7**	
Frecuencia observada	15	26	24	18	$n = 83$
Frecuencia esperada estimada	16.351	24.153	23.240	19.256	

Las cantidades de celda esperadas estimadas son $83\pi_i(\hat{\theta})$, donde $\hat{\theta}$ es el valor de θ que maximiza

$$\begin{aligned}
&\{\theta^4 + (1-\theta)^4\}^{15}\cdot\{4\theta(1-\theta)[\theta^3 + (1-\theta)^3]\}^{26} \\
&\qquad\cdot\{10\theta^2(1-\theta)^2[\theta^2 + (1-\theta)^2]\}^{24}\cdot\{20\theta^3(1-\theta)^3\}^{18}
\end{aligned} \qquad (14.5)$$

Los métodos estándar de cálculo no dan una fórmula sencilla para el valor que lleva al máximo a $\hat{\theta}$, de modo que debe calcularse usando métodos numéricos. El resultado es $\hat{\theta} = .654$, del que se calculan $\pi_i(\hat{\theta})$ y las cantidades de celda esperadas estimadas. El valor calculado de χ^2 es .360, y (como $k - 1 - m = 4 - 1 - 1 = 2$) $\chi^2_{.10,2} = 4.605$. Por tanto, no hay razón para rechazar el modelo simple como se aplica a la serie de juegos para decidir el campeonato de la NHL.

El artículo citado también consideró información de la Serie Mundial (de beisbol) para el periodo 1903-1973. Para el modelo simple, $\chi^2 = 5.97$, de modo que el modelo no parece apropiado. La razón sugerida para esto es que para el modelo simple

$$P(\text{la serie dura 6 juegos} \mid \text{la serie dura al menos 6 juegos}) \geq .5 \qquad (14.6)$$

mientras que, de las 38 series que en realidad duraron al menos seis juegos, sólo 13 duraron exactamente seis. Se introduce entonces el siguiente modelo alternativo:

$$\begin{aligned}
\pi_1(\theta_1, \theta_2) &= \theta_1^4 + (1-\theta_1)^4 \\
\pi_2(\theta_1, \theta_2) &= 4\theta_1(1-\theta_1)[\theta_1^3 + (1-\theta_1)^3] \\
\pi_3(\theta_1, \theta_2) &= 10\theta_1^2(1-\theta_1)^2\theta_2 \\
\pi_4(\theta_1, \theta_2) &= 10\theta_1^2(1-\theta_1)^2(1-\theta_2)
\end{aligned}$$

Las primeras dos π_i son idénticas al modelo simple, mientras que θ_2 es la probabilidad condicional de (14.6) (que ahora puede ser cualquier número entre 0 y 1). Los valores de $\hat{\theta}_1$ y $\hat{\theta}_2$ que maximizan la expresión análoga a la expresión (14.5) se determinan numéricamente como $\hat{\theta}_1 = .614$, $\hat{\theta}_2 = .342$. En la tabla 14.6 aparece un resumen, y $\chi^2 = .384$. Puesto que se estiman dos parámetros, el grado de libertad es igual a $k - 1 - m = 1$ con $\chi^2_{.10,1} = 2.706$, lo cual indica un buen ajuste de la información al nuevo modelo.

Tabla 14.6 Cantidades observadas y esperadas para el modelo más complejo

Número de juegos jugados	**4**	**5**	**6**	**7**
Frecuencia observada	12	16	13	25
Frecuencia esperada estimada	10.85	18.08	12.68	24.39

Una de las condiciones en las θ_i del teorema es que son funcionalmente independientes una de otra. Esto es, ningún θ_i se puede determinar a partir de los valores de otros θ_i, de modo que m es el número de parámetros estimados funcionalmente independientes. Una regla empírica general para los grados de libertad en una prueba chi cuadrada es la siguiente.

$$\chi^2 \text{ gl} = \begin{pmatrix}\text{número de cantidades de}\\ \text{celda determinadas libremente}\end{pmatrix} - \begin{pmatrix}\text{número de parámetros}\\ \text{independientes estimados}\end{pmatrix}$$

Esta regla se usará en relación con diversas pruebas de chi cuadrada diferentes en la siguiente sección.

Bondad de ajuste para distribuciones discretas

Numerosos experimentos comprenden la observación de una muestra aleatoria $X_1, X_2, \ldots, X_n$ de alguna distribución discreta. Entonces se puede desear investigar si la distribución básica es miembro de una familia particular, por ejemplo la familia Poisson o negativa binomial. En el caso de una distribución Poisson y una negativa binomial, el conjunto de posibles valores es infinito, de modo que los valores deben agruparse en k subconjuntos antes que pueda usarse una prueba chi cuadrada. Las agrupaciones deben hacerse de modo que la frecuencia esperada en cada celda (grupo) sea al menos 5. La última celda corresponderá entonces a X valores de $c, c + 1, c + 2, \ldots$ para algún valor de c.

Esta agrupación puede complicar de manera considerable el cálculo de los $\hat{\theta}_i$ y cantidades de celda esperadas estimadas. Esto se debe a que el teorema exige que las $\hat{\theta}_i$ se obtengan de las cantidades de celda $N_1, \ldots, N_k$ más que los valores muestrales $X_1, \ldots, X_n$.

Ejemplo 14.8 La tabla 14.7 presenta información de cantidades sobre el número de plantas *Larrea divaricata* halladas en cada uno de los 48 cuadrantes de muestreo, como se publica en el artículo "Some Sampling Characteristics of Plants and Arthropods of the Arizona Desert" (*Ecology*, 1962: 567–571).

Tabla 14.7 Cantidades observadas para el ejemplo 14.8

Celda	**1**	**2**	**3**	**4**	**5**
Número de plantas	**0**	**1**	**2**	**3**	**≥4**
Frecuencia	9	9	10	14	6

El autor del artículo ajustó una distribución Poisson a los datos. Denote con μ el parámetro Poisson y suponga por el momento que las seis cantidades de la celda 5 eran en realidad 4, 4, 5, 5, 6, 6. Entonces, denotando los valores muestrales por $x_1, \ldots, x_{48}$, nueve de las x_i fueron 0, nueve fueron 1, y así sucesivamente. La probabilidad de la muestra observada es

$$\frac{e^{-\mu}\mu^{x_1}}{x_1!} \cdot \cdots \cdot \frac{e^{-\mu}\mu^{x_{48}}}{x_{48}!} = \frac{e^{-48\mu}\mu^{\Sigma x_i}}{x_1! \cdot \cdots \cdot x_{48}!} = \frac{e^{-48\mu}\mu^{101}}{x_1! \cdot \cdots \cdot x_{48}!}$$

El valor de μ para el que esto se maximiza es $\hat{\mu} = \sum x_i/n = 101/48 = 2.10$ (el valor publicado en el artículo).

No obstante, la $\hat{\mu}$ requerida para χ^2 se obtiene de maximizar la expresión (14.4) más que la probabilidad de la muestra completa. Las probabilidades de celda son

$$\pi_i(\mu) = \frac{e^{-\mu}\mu^{i-1}}{(i-1)!} \quad i = 1, 2, 3, 4$$

$$\pi_5(\mu) = 1 - \sum_{i=0}^{3} \frac{e^{-\mu}\mu^i}{i!}$$

de modo que el lado derecho de (14.4) se convierte en

$$\left[\frac{e^{-\mu}\mu^0}{0!}\right]^9 \left[\frac{e^{-\mu}\mu^1}{1!}\right]^9 \left[\frac{e^{-\mu}\mu^2}{2!}\right]^{10} \left[\frac{e^{-\mu}\mu^3}{3!}\right]^{14} \left[1 - \sum_{i=0}^{3} \frac{e^{-\mu}\mu^i}{i!}\right]^6$$

No hay una fórmula fácil para $\hat{\mu}$, el valor de μ para maximizar, en esta última expresión, por lo que debe obtenerse numéricamente. ■

Debido a que las estimaciones del parámetro suelen ser mucho más difíciles de calcular del grupo de datos que de la muestra completa, prácticamente siempre se calculan usando este último método. Cuando se usan estos estimadores "completos" en el estadístico chi cuadrada, la distribución del estadístico se altera y una prueba α de nivel ya no es especificada por el valor crítico $\chi^2_{\alpha,k-1-m}$.

TEOREMA

Sean $\hat{\theta}_1, \ldots, \hat{\theta}_m$ los máximos estimadores de probabilidad de $\theta_1, \ldots, \theta_m$ con base en la muestra completa $X_1, \ldots, X_n$, y denótese con χ^2 el estadístico basado en estos estimadores. Entonces el valor crítico c_α que especifica una prueba de nivel α de cola superior satisface a

$$\chi^2_{\alpha,k-1-m} \leq c_\alpha \leq \chi^2_{\alpha,k-1} \tag{14.7}$$

El procedimiento de prueba implicado por este teorema es el siguiente:

Si $\chi^2 \geq \chi^2_{\alpha,k-1}$, rechace H_0.
Si $\chi^2 \leq \chi^2_{\alpha,k-1-m}$, no rechace H_0. (14.8)
Si $\chi^2_{\alpha,k-1-m} < \chi^2 < \chi^2_{\alpha,k-1}$, sin decisión.

Ejemplo 14.9 (Continuación del ejemplo 14.8)

Con el uso de $\hat{\mu} = 2.10$, las cantidades de celda esperadas estimadas se calculan de $n\pi_i(\hat{\mu})$, donde $n = 48$. Por ejemplo,

$$n\pi_1(\hat{\mu}) = 48 \cdot \frac{e^{-2.1}(2.1)^0}{0!} = (48)(e^{-2.1}) = 5.88$$

Del mismo modo, $n\pi_2(\hat{\mu}) = 12.34$, $n\pi_3(\hat{\mu}) = 12.96$, $n\pi_4(\hat{\mu}) = 9.07$, y $n\pi_5(\mu) = 48 - 5.88 - \cdots - 9.07 = 7.75$. Entonces

$$\chi^2 = \frac{(9 - 5.88)^2}{5.88} + \cdots + \frac{(6 - 7.75)^2}{7.75} = 6.31$$

Como $m = 1$ y $k = 5$, al nivel .05 se necesita $\chi^2_{.05,3} = 7.815$ y $\chi^2_{.05,4} = 9.488$. Como $6.31 \leq 7.815$, no se rechaza H_0; al nivel de 5%, la distribución Poisson da un ajuste razonable a los datos. Nótese que $\chi^2_{.10,3} = 6.251$ y $\chi^2_{.10,4} = 7.779$, de modo que al nivel .10 se tendría que retener un juicio sobre si la distribución Poisson era apropiada. ■

A veces hasta las estimaciones de máxima probabilidad basadas en la muestra completa son muy difíciles de calcular. Éste es el caso, por ejemplo, para la distribución binomial negativa (generalizada) de dos parámetros. En tales situaciones, se usan con frecuencia estimaciones del método de momentos y las χ^2 resultantes se comparan con $\chi^2_{\alpha,k-1-m}$, aunque no se sabe hasta qué punto el uso de estimadores de momento afecta al verdadero valor crítico.

Bondad de ajuste para distribuciones continuas

La prueba chi cuadrada también se puede usar para probar si la muestra proviene de una familia especificada de distribuciones continuas, como es el caso de la familia exponencial o la familia normal. La preferencia de celdas (intervalos de clase) es todavía más arbitraria en el caso continuo que en el discreto. Para asegurar que la prueba chi cuadrada es válida, las celdas deben escogerse independientemente de las observaciones muestrales. Una vez escogidas las celdas, casi siempre es muy difícil estimar parámetros no especificados (por ejemplo μ y σ en el caso normal) a partir de las cantidades de celda observadas, de modo que en su lugar se calculan estimadores de máxima probabilidad basados en la muestra completa. El valor crítico c_α de nuevo satisface a (14.7) y el procedimiento de prueba está dado por (14.8).

Ejemplo 14.10

El Institute of Nutrition of Central America and Panama (INCAP) ha llevado a cabo extensos estudios de dietas y proyectos de investigación en Centroamérica. En un estudio que apareció en la edición de noviembre de 1964 del *American Journal of Clinical Nutrition* ("The Blood Viscosity of Various Socioeconomic Groups in Guatemala"), se publicaron las mediciones de colesterol seroso total de una muestra de 49 indígenas rurales de bajos ingresos, como sigue (en mg/L):

204	108	140	152	158	129	175	146	157	174	192	194	144	152	135	223	145
231	115	131	129	142	114	173	226	155	166	220	180	172	143	148	171	143
124	158	144	108	189	136	136	197	131	95	139	181	165	142	162		

¿Es posible que el nivel de colesterol seroso esté normalmente distribuido para esta población? Suponga que, antes de la muestra, se pensaba que los posibles valores para μ y σ eran 150 y 30, respectivamente. Los siete intervalos de clase igualmente probables para la distribución estándar normal son $(-\infty, -1.07)$, $(-1.07, -.57)$, $(-.57, -.18)$, $(-.18, .18)$, $(.18, .57)$, $(.57, 1.07)$ y $(1.07, \infty)$, con cada punto extremo proporcionando también la distancia en desviaciones estándar desde la media para cualquier otra distribución normal.

Para $\mu = 150$ y $\sigma = 30$, estos intervalos se convierten en $(-\infty, 117.9)$, $(117.9, 132.9)$, $(132.9, 144.6)$, $(144.6, 155.4)$, $(155.4, 167.1)$, $(167.1, 182.1)$, y $(182.1, \infty)$.

Para obtener las probabilidades estimadas de celda $\pi_1(\hat{\mu}, \hat{\sigma}), \ldots, \pi_7(\hat{\mu}, \hat{\sigma})$, primero se necesitan los estimadores de máxima probabilidad $\hat{\mu}$ y $\hat{\sigma}$. En el capítulo 6, la estimación de máxima probabilidad de σ resultó ser $[\sum(x_i - \bar{x})^2/n]^{1/2}$ (más que s), de modo que $s = 31.75$,

$$\hat{\mu} = \bar{x} = 157.02 \qquad \hat{\sigma} = \left[\frac{\sum(x_i - \bar{x})^2}{n}\right]^{1/2} = \left[\frac{(n-1)s^2}{n}\right]^{1/2} = 31.42$$

Cada $\pi_i(\hat{\mu}, \hat{\sigma})$ es entonces la probabilidad de que una variable aleatoria normal X con media 157.02 y desviación estándar 31.42 caiga en el intervalo de i-ésima clase. Por ejemplo,

$$\pi_2(\hat{\mu}, \hat{\sigma}) = P(117.9 \le X \le 132.9) = P(-1.25 \le Z \le -.77) = .1150$$

así $n\pi_2(\hat{\mu}, \hat{\sigma}) = 49(.1150) = 5.64$. Las cantidades de celda observadas y esperadas se muestran en la tabla 14.8.

La χ^2 calculada es 4.60. Con $k = 7$ celdas y $m = 2$ parámetros estimados, $\chi^2_{.05,k-1} = \chi^2_{.05,6} = 12.592$ y $\chi^2_{.05,k-1-m} = \chi^2_{.05,4} = 9.488$. Como $4.60 \le 9.488$, una distribución normal da un muy buen ajuste a los datos.

Tabla 14.8 Cantidades observadas y esperadas para el ejemplo 14.10

Celda	**$(-\infty, 117.9)$**	**(117.9, 132.9)**	**(132.9, 144.6)**	**(144.6, 155.4)**
Observada	5	5	11	6
Esperada estimada	5.17	5.64	6.08	6.64

Celda	**(155.4, 167.1)**	**(167.1, 182.1)**	**(182.1,∞)**
Observada	6	7	9
Esperada estimada	7.12	7.97	10.38

Ejemplo 14.11 El artículo "Some Studies on Tuft Weight Distribution in the Opening Room" (*Textile Research J.*, 1976: 567–573) publica los datos siguientes sobre la distribución del peso X en el peso de mechones de salida (mg) de fibras de algodón para el peso de entrada $x_0 = 70$.

Intervalo	**0–8**	**8–16**	**16–24**	**24–32**	**32–40**	**40–48**	**48–56**	**56–64**	**64–70**
Frecuencia observada	20	8	7	1	2	1	0	1	0
Frecuencia esperada	18.0	9.9	5.5	3.0	1.8	.9	.5	.3	.1

Los autores postularon una distribución exponencial truncada:

$$H_0: f(x) = \frac{\lambda e^{-\lambda x}}{1 - e^{-\lambda x_0}} \qquad 0 \le x \le x_0$$

La media de esta distribución es

$$\mu = \int_0^{x_0} x f(x)\, dx = \frac{1}{\lambda} - \frac{x_0 e^{-\lambda x_0}}{1 - e^{-\lambda x_0}}$$

El parámetro λ se estimó al sustituir μ con $\bar{x} = 13.086$ y resolver la ecuación resultante para obtener $\hat{\lambda} = .0742$ (de modo que $\hat{\lambda}$ es una estimación del método de momentos y no

un estimador de máxima probabilidad). Entonces con $\hat{\lambda}$ sustituyendo a λ en $f(x)$, las frecuencias de celda esperadas estimadas como se exhibieron antes se calculan como

$$40\pi_i(\hat{\lambda}) = 40P(a_{i-1} \le X < a_i) = 40\int_{a_{i-1}}^{a_i} f(x)\,dx = \frac{40(e^{-\hat{\lambda}a_{i-1}} - e^{-\hat{\lambda}a_i})}{1 - e^{-\hat{\lambda}x_o}}$$

donde $[a_{i-1}, a_i)$ es el i-ésimo intervalo de clase. Para obtener cantidades de celda esperadas de al menos 5, las últimas seis celdas se combinan para dar cantidades observadas de 20, 8, 7, 5 y cantidades esperadas de 18.0, 9.9, 5.5, 6.6. El valor calculado de chi cuadrada es entonces $\chi^2 = 1.34$. Como $\chi^2_{.05,2} = 5.992$, H_0 no es rechazada, y el modelo exponencial truncado da un buen ajuste. ■

Una prueba especial para normalidad

Las gráficas de probabilidad se introdujeron en la sección 4.6, como un método informal para evaluar la posibilidad de que cualquier distribución poblacional especificada sea aquella de la que se seleccionó la muestra dada. Cuanto más recta sea la gráfica de probabilidad, más posible es la distribución en la que está basada la gráfica. Una gráfica de probabilidad normal se emplea para verificar si *cualquier* miembro de la familia de distribución normal es posible. Denótese las x_i muestrales cuando sean ordenadas de menor a mayor por $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$. Entonces la gráfica sugerida para verificar normalidad fue una gráfica de los puntos $(x_{(i)}, y_i)$, donde $y_i = \Phi^{-1}((i - .5)/n)$.

Una medida cuantitativa de la magnitud a la que los puntos se agrupan alrededor de una recta es el coeficiente r de correlación muestral introducido en el capítulo 12. Considere calcular r para los n pares $(x_{(1)}, y_1), \ldots, (x_{(n)}, y_n)$. Las y_i aquí no son valores observados en una muestra aleatoria de una población y, de modo que las propiedades de esta r son muy diferentes a las descritas en la sección 12.5. No obstante, es cierto que cuanto más se desvíe r de 1, menor es la probabilidad de que la gráfica se asemeje a una recta (recuerde que una gráfica de probabilidad debe tener pendiente ascendente). Esta idea también se puede extender para obtener un procedimiento formal de prueba: rechazar la hipótesis de normalidad poblacional si $r \le c_\alpha$, donde c_α es un valor crítico seleccionado para obtener el nivel α deseado de significación. Esto es, el valor crítico se selecciona de modo que cuando la distribución poblacional sea normal en realidad, la probabilidad de obtener un valor r que sea a lo sumo c_α (y así rechazar incorrectamente H_0) es la α deseada. Los creadores del paquete estadístico computacional Minitab dan valores críticos para $\alpha = .10$, .05, y .01 en combinación con tamaños muestrales diferentes. Estos valores críticos están basados en una definición ligeramente diferente de las y_i que las dadas antes.

Minitab también construirá una gráfica de probabilidad normal con estas y_i. La gráfica tendrá un aspecto casi idéntico al basado en las y_i previas. Cuando haya varias $x_{(i)}$ empatadas, Minitab calcula r usando el promedio de las y_i correspondientes como el segundo número de cada par.

Sea $y_i = \Phi^{-1}[(i - .375)/(n + .25)]$, y calcule el coeficiente r de correlación muestral para los n pares $(x_{(1)}, y_1), \ldots, (x_{(n)}, y_n)$. La **prueba Ryan-Joiner** de

H_0: la distribución poblacional es normal

contra

H_a: la distribución poblacional no es normal

consiste en rechazar H_0 cuando $r \le c_\alpha$. Los valores críticos c_α se dan en la tabla A.12 del apéndice para diversos niveles α de significación y tamaños muestrales n.

Ejemplo 14.12 La siguiente muestra de $n = 20$ observaciones sobre voltaje de ruptura dieléctrico de una pieza de resina epóxica apareció primero en el ejemplo 4.30.

y_i	–1.871	–1.404	–1.127	–.917	–.742	–.587	–.446	–.313	–.186	–.062
$x_{(i)}$	24.46	25.61	26.25	26.42	26.66	27.15	27.31	27.54	27.74	27.94
y_i	.062	.186	.313	.446	.587	.742	.917	1.127	1.404	1.871
$x_{(i)}$	27.98	28.04	28.28	28.49	28.50	28.87	29.11	29.13	29.50	30.88

Se pidió a Minitab que realizara la prueba de Ryan-Joiner y el resultado aparece en la figura 14.3. El valor del estadístico de prueba es $r = .9881$, y la tabla A.12 del apéndice da .9600 como el valor crítico que captura el área de cola inferior .10, bajo la curva de distribución muestral r cuando $n = 20$ y la distribución básica en realidad es normal. Como $.9881 > .9600$, la hipótesis nula de normalidad no puede ser rechazada incluso para un nivel de significancia de hasta .10.

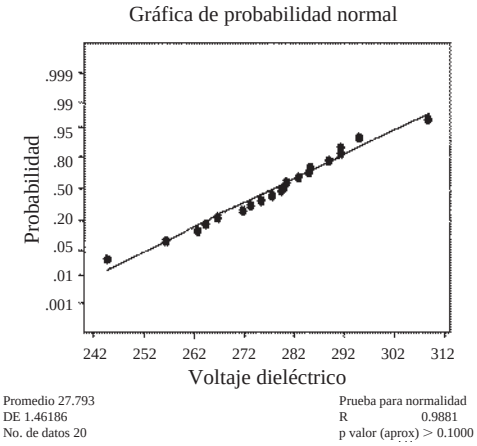


Figura 14.3 Salida de Minitab de la prueba Ryan-Joiner para los datos del ejemplo 14.12

EJERCICIOS Sección 14.2 (12–23)

12. Considere una gran población de familias en las que cada una tiene exactamente tres hijos. Si los géneros de los tres hijos de cualquier familia son independientes entre sí, el número de hijos hombres de una familia seleccionada al azar tendrá una distribución binomial basada en tres intentos.

a. Suponga que una muestra aleatoria de 160 familias da los resultados siguientes. Pruebe las hipótesis relevantes procediendo como en el ejemplo 14.5.

Número de hijos hombres	0	1	2	3
Frecuencia	14	66	64	16

b. Suponga que una muestra aleatoria de familias de una población que no es de seres humanos dio como resultado frecuencias observadas de 15, 20, 12 y 3, respectivamente. ¿La prueba chi cuadrada estaría basada en el mismo número de grados de libertad que la prueba del inciso (a)? Explique.

13. Un estudio de esterilidad en moscas de la fruta ("Hybrid Dysgenesis in *Drosophila melanogaster:* The Biology of Female and Male Sterility", *Genetics*, 1979: 161–174) publica los datos siguientes sobre el número de ovarios desarrollados por cada mosca hembra en una muestra de 1388. Un modelo de esterilidad unilateral expresa que cada ovario se desarrolla con alguna probabilidad p, independientemente del otro ovario. Pruebe el ajuste de este modelo usando χ^2.

x = número de ovarios desarrollados	0	1	2
Cantidad observada	1212	118	58

14. El artículo "Feeding Ecology of the Red-Eyed Vireo and Associated Foliage-Gleaning Birds" (*Ecological Monographs*, 1971: 129–152) presenta los datos siguientes sobre la variable X = número de saltos antes del primer vuelo y precedido de un vuelo. El autor propuso y ajustó entonces una distribución de

probabilidad geométrica [$p(x) = P(X = x) = p^{x-1} \cdot q$ para $x = 1, 2, \ldots$, donde $q = 1 - p$] a los datos. El tamaño total muestral fue $n = 130$.

x	1	2	3	4	5	6	7	8	9	10	11	12
Número de veces que se observa x	48	31	20	9	6	5	4	2	1	1	2	1

a. La probabilidad es $(p^{x_1-1} \cdot q) \cdot \cdots \cdot (p^{x_n-1} \cdot q) = p^{\sum x_i - n} \cdot q^n$. Demuestre que los estimadores de máxima probabilidad de p están dados por $\hat{p} = (\sum x_i - n)/\sum x_i$, y calcule $\hat{p}$ para la información dada.

b. Estime las cantidades esperadas de celda usando $\hat{p}$ del inciso (a) [cantidades esperadas de celda $= n \cdot (\hat{p})^{x-1} \cdot \hat{q}$ para $x = 1, 2, \ldots$], y pruebe el ajuste del modelo usando una prueba χ^2 al combinar las cantidades para $x = 7, 8, \ldots$, y 12 en una celda ($x \geq 7$).

15. Cierto tipo de linterna eléctrica se vende con las cuatro baterías incluidas. Se obtiene una muestra aleatoria de 150 linternas y se determina el número de baterías defectuosas; el resultado son los datos siguientes:

Número defectuosas	0	1	2	3	4
Frecuencia	26	51	47	16	10

Sea X el número de baterías defectuosas de una linterna seleccionada al azar. Pruebe la hipótesis nula de que la distribución de X es Bin(4, θ). Esto es, con $p_i = P(i$ defectuosas), pruebe

$$H_0: p_i = \binom{4}{i} \theta^i (1 - \theta)^{4-i} \qquad i = 0, 1, 2, 3, 4$$

[*Sugerencia:* para obtener los estimadores de máxima probabilidad de θ, escriba la probabilidad (la función a ser maximizada) como $\theta^u(1 - \theta)^v$, donde los exponentes u y v son funciones lineales de las cantidades de celdas. Luego tome el logaritmo natural, derive con respecto a θ, iguale a cero el resultado y despeje $\hat{\theta}$.]

16. En un experimento de genética, unos investigadores observaron 300 cromosomas de un tipo particular y contaron el número de intercambios de hermana cromátida en cada uno ("On the Nature of Sister-Chromatid Exchanges in 5-Bromodeoxyuridine-Substituted Chromosomes", *Genetics*, 1979: 1251–1264). Se teorizó un modelo Poisson para la distribución del número de intercambios. Pruebe el ajuste de una distribución Poisson a los datos estimando primero μ y luego combinando las cantidades para $x = 8$ y $x = 9$ en una celda.

x = número de intercambios	0	1	2	3	4	5	6	7	8	9
Cantidades observadas	6	24	42	59	62	44	41	14	6	2

17. Un artículo en *Annals of Mathematical Statistics* publica los siguientes datos sobre el número de taladradores en cada uno de los 120 grupos de taladradores. ¿La función de masa de probabilidad de Poisson da un modelo posible para la distribución del número de taladradores en un grupo? [*Sugerencia:* sume las frecuencias para 7, 8, . . . , 12 para establecer una sola categoría "≥ 7".]

Número de taladradores	0	1	2	3	4	5	6	7	8	9	10	11	12
Frecuencia	24	16	16	18	15	9	6	5	3	4	3	0	1

18. El artículo "A Probabilistic Analysis of Dissolved Oxygen–Biochemical Oxygen Demand Relationship in Streams" (*J. Water Resources Control Fed.*, 1969: 73–90) publica datos sobre la rapidez de oxigenación en arroyos a 20°C en cierta región. La media muestral y desviación estándar se calcularon como $\bar{x} = .173$ y $s = .066$, respectivamente. Con base en la distribución de frecuencia siguiente, ¿puede concluirse que la rapidez de oxigenación es una variable normalmente distribuida? Use la prueba chi cuadrada con $\alpha = .05$.

Rapidez (por día)	**Frecuencia**
Abajo de .100	12
.100–abajo de .150	20
.150–abajo de .200	23
.200–abajo de .250	15
.250 o más	13

19. Cada faro delantero de un auto sometido a inspección anual puede enfocarse ya sea demasiado alto (H), demasiado bajo (L), o bien (N). La verificación de los dos faros simultáneamente (y sin distinguir entre izquierdo y derecho) da los seis posibles resultados *HH*, *LL*, *NN*, *HL*, *HN* y *LN*. Si las posibilidades (proporciones poblacionales) para la dirección de enfoque de un solo faro son $P(H) = \theta_1$, $P(L) = \theta_2$, y $P(N) = 1 - \theta_1 - \theta_2$, y si los dos faros se enfocan de modo independiente uno del otro, las probabilidades de los seis resultados para un auto seleccionado al azar son las siguientes:

$$p_1 = \theta_1^2 \quad p_2 = \theta_2^2 \quad p_3 = (1 - \theta_1 - \theta_2)^2$$
$$p_4 = 2\theta_1\theta_2 \quad p_5 = 2\theta_1(1 - \theta_1 - \theta_2)$$
$$p_6 = 2\theta_2(1 - \theta_1 - \theta_2)$$

Use los datos siguientes para probar la hipótesis nula

$$H_0: p_1 = \pi_1(\theta_1, \theta_2), \ldots, p_6 = \pi_6(\theta_1, \theta_2)$$

donde las $\pi_i(\theta_1, \theta_2)$ se dieron previamente.

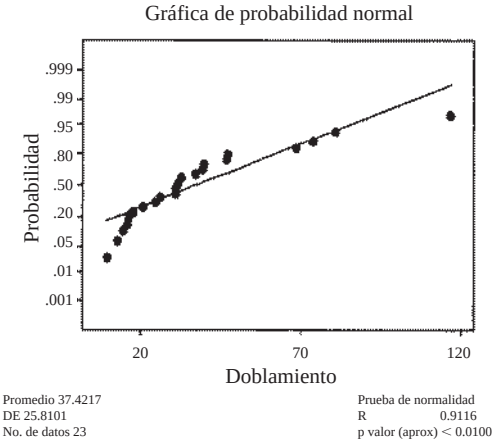
Resultado	*HH*	*LL*	*NN*	*HL*	*HN*	*LN*
Frecuencia	49	26	14	20	53	38

[*Sugerencia:* escriba la probabilidad como una función de θ_1 y θ_2, tome el logaritmo natural, luego calcule $\partial/\partial\theta_1$ y $\partial/\partial\theta_2$, iguálelas a cero y despeje $\hat{\theta}_1, \hat{\theta}_2$.]

20. El artículo "Compatibility of Outer and Fusible Interlining Fabrics in Tailored Garments" (*Textile Res. J.*, 1997: 137–142)

dio las siguientes observaciones en rigidez al doblamiento ($\mu N \cdot m$) para especímenes de telas de mediana calidad, de las cuales se obtuvo la salida Minitab siguiente:

24.6	12.7	14.4	30.6	16.1	9.5	31.5	17.2
46.9	68.3	30.8	116.7	39.5	73.8	80.6	20.3
25.8	30.9	39.2	36.8	46.6	15.6	32.3	



¿Usaría el lector un intervalo de confianza *t* de una muestra para estimar un promedio verdadero de rigidez al doblamiento? Explique su razonamiento.

21. El artículo del que se obtuvo la información del ejercicio 20 también dio los datos siguientes sobre la proporción de masa compuesta/masa de tela exterior para especímenes de tela de alta calidad.

1.15	1.40	1.34	1.29	1.36	1.26	1.22	1.40
1.29	1.41	1.32	1.34	1.26	1.36	1.36	1.30
1.28	1.45	1.29	1.28	1.38	1.55	1.46	1.32

Minitab dio $r = .9852$ como el valor del estadístico de prueba Ryan-Joiner e indicó que el valor *P* es $> .10$. ¿El lector utilizaría la prueba *t* de una muestra para probar las hipótesis acerca del valor del verdadero promedio de la relación? ¿Por qué sí o por qué no?

22. El artículo "A Method for the Estimation of Alcohol in Fortified Wines Using Hydrometer Baumé and Refractometer Brix" (*Amer. J. of Enol. and Vitic.*, 2006: 486–490) dio las mediciones duplicadas sobre el contenido de alcohol destilado (%) para una muestra de 35 vinos de Oporto. Éstos son los promedios de las medidas por duplicado:

15.30	16.20	16.35	17.15	17.48	17.73	17.75	17.85	18.00
18.68	18.82	18.85	19.03	19.07	19.08	19.17	19.20	19.20
19.33	19.37	19.45	19.48	19.50	19.58	19.60	19.62	19.90
19.97	20.00	20.05	21.22	22.25	22.75	23.25	23.78	

Use la prueba de Ryan-Joiner para decidir un nivel de significancia de .05 si una distribución normal proporciona un modelo plausible de contenido de alcohol.

23. El artículo "Nonbloated Burned Clay Aggregate Concrete" (*J. of Materials*, 1972: 555–563) publica los siguientes datos sobre resistencia flexional de 7 días de muestras de concreto con agregado de arcilla quemada sin curar (en libras por pulgada cuadrada):

257	327	317	300	340	340	343	374	377	386
383	393	407	407	434	427	440	407	450	440
456	460	456	476	480	490	497	526	546	700

Pruebe al nivel .10 para determinar si la resistencia flexional es una variable distribuida normalmente.

14.3 Tablas de contingencia mutuas (o bidireccionales)

En los escenarios de las secciones 14.1 y 14.2, las frecuencias observadas fueron mostradas en un solo renglón dentro de una tabla rectangular. Ahora se estudian problemas en los que los datos también están formados por cantidades o frecuencias, pero la tabla de información ahora tendrá *I* filas ($I \geq 2$) y *J* columnas, por tanto *IJ* celdas. Hay dos situaciones que por lo general se encuentran y en las que se muestran los datos:

1. Hay *I* poblaciones de interés, cada una correspondiente a una fila diferente de la tabla y cada población está dividida en las mismas *J* categorías. Se toma una muestra de la *i*-ésima población ($i = 1, \ldots, I$) y las cantidades se introducen en las celdas de la *i*-ésima fila de la tabla. Por ejemplo, los clientes de cada una de las $I = 3$ cadenas de tiendas departamentales podrían disponer de las mismas $J = 5$ categorías de pago: contado, cheque, tarjeta de crédito de la tienda, Visa y MasterCard.
2. Hay una sola población de interés, con cada individuo de la población clasificado con respecto a dos factores diferentes. Hay *I* categorías asociadas con el primer factor, y *J* categorías asociadas con el segundo factor. Se toma una sola muestra y el número de individuos pertenecientes tanto a la categoría *i* del factor 1 como a la categoría *j* del factor 2 se introduce en la celda de la fila *i*, columna *j* ($i = 1, \ldots, I$; $j = 1, \ldots, J$). Como ejemplo, los clientes que hagan una compra podrían clasificarse de acuerdo con el

departamento en el que hicieron la compra, con $I = 6$ departamentos, y de acuerdo con la forma de pago, con $J = 5$ como en (1) líneas antes.

Denótese con n_{ij} el número de individuos de la(s) muestra(s) que caen en la (i, j)-ésima celda (fila i, columna j) de la tabla; es decir, la (i, j)-ésima cantidad de celda. La tabla que presenta las n_{ij} se denomina **tabla de contingencia mutua;** un prototipo se muestra en la tabla 14.9.

Tabla 14.9 Una tabla de contingencia mutua

	1	**2**	**. . .**	***j***	**. . .**	***J***
1	n_{11}	n_{12}	. . .	n_{1j}	. . .	n_{1J}
2	n_{21}					⋮
⋮	⋮					
i	n_{i1}	. . .		n_{ij}	. . .	
⋮	⋮					
I	n_{I1}	. . .				n_{IJ}

En situaciones del tipo 1, se desea investigar si las proporciones de las diferentes categorías son iguales para todas las poblaciones. La hipótesis nula expresa que las poblaciones son *homogéneas* con respecto a estas categorías. En situaciones del tipo 2, se investiga si las categorías de los dos factores se presentan independientemente una de otra en la población.

Prueba de homogeneidad

Se supone que cada uno de los individuos de las poblaciones I pertenece exactamente a una de las J categorías. Una muestra de n_i individuos se toma de la i-ésima población; sea $n = \sum n_i$ y

$n_{ij} =$ número de individuos de la i-ésima muestra que caen en la categoría j

$n_{\cdot j} = \sum_{i=1}^{I} n_{ij} =$ número total de individuos entre la muestra n que cae en la categoría j

Las n_{ij} se registran en una tabla de contingencia mutua con I filas y J columnas. La suma de las n_{ij} de la i-ésima fila es n_i, mientras que la suma de entradas de la j-ésima columna es $n_{\cdot j}$.

Sea

$$p_{ij} = \text{proporción de individuos de la población } i \text{ que cae en la categoría } j$$

Así, para la población 1, las proporciones J son $p_{11}, p_{12}, \ldots, p_{1J}$ (que suman 1) y análogamente para otras poblaciones. La **hipótesis nula de homogeneidad** expresa que la proporción de individuos de la categoría j es la misma para cada población y que esto es cierto para toda categoría, es decir, j, $p_{1j} = p_{2j} = \cdots = p_{Ij}$.

Cuando H_0 es verdadera, se puede usar $p_1, p_2, \ldots, p_J$ para denotar las proporciones poblacionales de las J categorías diferentes; estas proporciones son comunes para todas las poblaciones I. El número esperado de individuos en la i-ésima muestra que cae en la j-ésima categoría cuando H_0 es verdadera es entonces $E(N_{ij}) = n_i \cdot p_j$. Para estimar $E(N_{ij})$, primero se debe estimar p_j, la proporción de la categoría j. Entre la muestra total de n individuos, $N_{\cdot j}$ cae en la categoría j, de modo que se usa $\hat{p}_j = N_{\cdot j}/n$ como el estimador (se puede demostrar que éste es el estimador de máxima probabilidad de p_j). La sustitución de la estimación $\hat{p}_j$ para p_j en $n_i p_j$ da una fórmula sencilla para cantidades esperadas estimadas bajo H_0:

$$\hat{e}_{ij} = \text{cantidad esperada estimada en celda } (i, j) = n_i \cdot \frac{n_{\cdot j}}{n}$$
$$= \frac{(\text{total de } i\text{-ésima fila})(\text{total de } j\text{-ésima columna})}{n} \qquad (14.9)$$

El estadístico de prueba también tiene la misma forma que en las situaciones del problema previo. El número de grados de libertad proviene de la regla empírica general. En cada fila de la tabla 14.9 hay $J - 1$ cantidades de celdas determinadas libremente (cada tamaño muestral n_i es fijo), de modo que hay un total de $I(J - 1)$ celdas determinadas libremente. Los parámetros $p_1, \ldots, p_J$ se estiman, pero como $\sum p_i = 1$, sólo $J - 1$ de éstos son independientes. Por lo tanto, el grado de libertad es gl $= I(J - 1) - (J - 1) = (J - 1)(I - 1)$.

Hipótesis nula: H_0: $p_{1j} = p_{2j} = \cdots = p_{Ij} \quad j = 1, 2, \ldots, J$

Hipótesis alternativa: H_a: H_0 no es verdadera

Valor de estadístico de prueba:

$$\chi^2 = \sum_{\text{todas las celdas}} \frac{(\text{observada} - \text{esperada estimada})^2}{\text{esperada estimada}} = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{(n_{ij} - \hat{e}_{ij})^2}{\hat{e}_{ij}}$$

Región de rechazo: $\chi^2 \geq \chi^2_{\alpha,(I-1)(J-1)}$

La información del valor P se puede obtener como se describe en la sección 14.1. La prueba se puede aplicar con seguridad mientras $\hat{e}_{ij} \geq 5$ para todas las celdas.

Ejemplo 14.13 Una compañía empaca un producto particular en latas de tres tamaños diferentes, cada uno con una línea de producción distinta. La mayor parte de las latas se apegan a especificaciones, pero un ingeniero de control de calidad ha identificado las siguientes razones de no cumplimiento:

1. Defecto en lata
2. Grieta en lata
3. Ubicación incorrecta de arillo
4. Arillo faltante
5. Otras

Se selecciona una muestra de unidades fuera de especificación de cada una de las tres líneas, y cada unidad se clasifica según la razón por la que está fuera de especificación; dio por resultado la siguiente información de tabla de contingencia:

		Razón para estar fuera de especificación					Tamaño muestral
		Defecto	Grieta	Ubicación	Faltante	Otras	
Línea de producción	**1**	34	65	17	21	13	150
	2	23	52	25	19	6	125
	3	32	28	16	14	10	100
	Total	89	145	58	54	29	375

¿Sugiere la información que las proporciones que caen en las diversas categorías fuera de especificación no son iguales para las tres líneas? Los parámetros de interés son las diversas proporciones, y las hipótesis relevantes son

H_0: las líneas de producción son homogéneas con respecto a las cinco categorías fuera de especificación; es decir, $p_{1j} = p_{2j} = p_{3j}$ para $j = 1, \ldots, 5$

H_a: las líneas de producción no son homogéneas con respecto a las categorías

Las frecuencias esperadas estimadas (suponiendo homogeneidad) deben calcularse ahora. Considere la primera categoría fuera de especificación para la primera línea de producción. Cuando las líneas son homogéneas,

$$\text{número esperado estimado entre las 150 unidades seleccionadas echadas a perder} = \frac{(\text{total primera fila})(\text{total primera columna})}{\text{total de tamaños muestrales}} = \frac{(150)(89)}{375} = 35.60$$

La contribución de la celda de la esquina superior izquierda a χ^2 es entonces

$$\frac{(\text{observada} - \text{esperada estimada})^2}{\text{esperada estimada}} = \frac{(34 - 35.60)^2}{35.60} = .072$$

Las otras contribuciones se calculan de manera semejante. La figura 14.4 muestra una salida Minitab para la prueba de chi cuadrada. La cantidad observada es el número de la parte superior de cada celda, y directamente debajo de ella está la cantidad esperada estimada. La contribución de cada celda a χ^2 aparece debajo de las cantidades, y el valor del estadístico de prueba es $\chi^2 = 14.159$. Todas las cantidades esperadas estimadas son al menos 5, de modo que no es necesario combinar categorías. La prueba está basada en $(3 - 1)(5 - 1) = 8$ grados de libertad. La tabla A.11 del apéndice muestra que los valores que capturan las áreas de cola superior de .08 y .075 bajo la curva de 8 grados de libertad son 14.06 y 14.26, respectivamente. Por tanto, el valor P está entre .075 y .08; Minitab da un valor $P = .079$. La hipótesis nula de homogeneidad no debe ser rechazada a los niveles de significación usuales de .05 y .01, pero debe ser rechazada para α mayor que .10.

```
Las cantidades esperadas se indican bajo los conteos observados
        defecto    grieta    ubicación    faltante     otro     Total
1            34        65           17          21       13       150
          35.60     58.00        23.20       21.60    11.60
2            23        52           25          19        6       125
          29.67     48.33        19.33       18.00     9.67
3            32        28           16          14       10       100
          23.73     38.67        15.47       14.40     7.73
Total        89       145           58          54       29       375

Chi cuadrada = 0.072 + 0.845 + 1.657 + 0.017 + 0.169 + 1.498 + 0.278 +
               1.661 + 0.056 + 1.391 + 2.879 + 2.943 + 0.018 + 0.011 +
               0.664 = 14.159

gl = 8, p = 0.079
```

Figura 14.4 Salida de Minitab para la prueba de chi cuadrada del ejemplo 14.13 ■

Prueba de independencia

Ahora el tema se concentra en la relación entre dos factores diferentes de una población individual. Cada individuo de la población pertenece a exactamente una de las I categorías, asociada con el primer factor y a exactamente una de las J categorías asociada con el segundo factor. Por ejemplo, la población de interés podría estar formada por todos los individuos que por lo regular ven por televisión las noticias nacionales, en el que el primer factor es una red preferida (ABC, CBS, NBC, o PBS, de modo que $I = 4$) y el segundo factor la filosofía política (liberal, moderado, o conservador, que da $J = 3$).

Para una muestra de n individuos tomada de la población, denote con n_{ij} el número entre los n que caen en la categoría i del primer factor y la categoría j del segundo factor. Las n_{ij} se pueden exhibir en una tabla de contingencia mutua con I filas y J columnas. En el caso de homogeneidad para I poblaciones, los totales de fila se fijaron por anticipado, y sólo los totales de la columna J fueron aleatorios. Ahora sólo el tamaño muestral total es fijo y las $n_{i\cdot}$ y $n_{\cdot j}$ son valores observados de variables aleatorias. Para expresar las hipótesis de interés, sea

$$p_{ij} = \text{la proporción de individuos de la población que pertenece a la categoría } i \text{ del factor 1 y a la categoría } j \text{ del factor 2}$$

$$= P(\text{un individuo seleccionado al azar cae en la categoría } i \text{ del factor 1 y en la categoría } j \text{ del factor 2})$$

Entonces

$$p_{i\cdot} = \sum_j p_{ij} = P(\text{un individuo seleccionado al azar cae en la categoría } i \text{ del factor 1})$$

$$p_{\cdot j} = \sum_i p_{ij} = P(\text{un individuo seleccionado al azar cae en la categoría } j \text{ del factor 2})$$

Recuerde que dos eventos A y B son independientes si $P(A \cap B) = P(A) \cdot P(B)$. *La hipótesis nula aquí dice que la categoría de un individuo con respecto al factor 1 es independiente de la categoría con respecto al factor 2*. En símbolos, esto se convierte en $p_{ij} = p_{i\cdot} \cdot p_{\cdot j}$ para todo par (i, j).

La cantidad esperada en la celda (i, j) es $n \cdot p_{ij}$, de modo que cuando la hipótesis nula es verdadera, $E(N_{ij}) = n \cdot p_{i\cdot} \cdot p_{\cdot j}$. Para obtener un estadístico de chi cuadrada, se deben estimar por tanto las $p_{i\cdot}(i = 1, \ldots, I)$ y las $p_{\cdot j}(j = 1, \ldots, J)$. Los estimadores de máxima probabilidad son

$$\hat{p}_{i\cdot} = \frac{n_{i\cdot}}{n} = \text{proporción muestral para la categoría } i \text{ del factor 1}$$

y

$$\hat{p}_{\cdot j} = \frac{n_{\cdot j}}{n} = \text{proporción muestral para la categoría } j \text{ del factor 2}$$

Esto da cantidades de celda esperadas estimadas idénticas a las del caso de homogeneidad.

$$\hat{e}_{ij} = n \cdot \hat{p}_{i\cdot} \cdot \hat{p}_{\cdot j} = n \cdot \frac{n_{i\cdot}}{n} \cdot \frac{n_{\cdot j}}{n} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}$$
$$= \frac{(\text{total de } i\text{-ésima fila})(\text{total de } j\text{-ésima columna})}{n}$$

El estadístico de prueba también es idéntico al empleado en pruebas de homogeneidad, como es el número de grados de libertad. Esto es porque el número de cantidades de celda determinadas libremente es $IJ - 1$, aunque sólo el n total se fija por anticipado. Hay I de $p_{i\cdot}$ estimadas, pero sólo $I - 1$ son estimadas de manera independiente porque $\sum p_{i\cdot} = 1$, y, análogamente $J - 1$ de las $p_{\cdot j}$ son estimadas de manera independiente, de modo que $I + J - 2$ parámetros se estiman independientemente. La regla empírica ahora da grados de libertad $= IJ - 1 - (I + J - 2) = IJ - I - J + 1 = (I - 1) \cdot (J - 1)$.

Hipótesis nula: $H_0: p_{ij} = p_{i\cdot} \cdot p_{\cdot j} \quad i = 1, \ldots, I; j = 1, \ldots, J$

Hipótesis alternativa: H_a: H_0 no es verdadera

Valor de estadístico de prueba:

$$\chi^2 = \sum_{\substack{\text{todas} \\ \text{las celdas}}} \frac{(\text{observada} - \text{esperada estimada})^2}{\text{esperada estimada}} = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{(n_{ij} - \hat{e}_{ij})^2}{\hat{e}_{ij}}$$

Región de rechazo: $\chi^2 \geq \chi^2_{\alpha,(I-1)(J-1)}$

De nuevo, la información del valor P se puede obtener como se describe en la sección 14.1. La prueba puede aplicarse con seguridad mientras $\hat{e}_{ij} \geq 5$ para todas las celdas.

Ejemplo 14.14 Un estudio de la relación entre las condiciones del equipo de gasolineras y la agresividad en los precios de gasolina ("An Analysis of Price Aggresiveness in Gasoline Marketing", *J. of Marketing Research*, 1970: 36–42) publica la siguiente información basada en una muestra de $n = 441$ gasolineras. Al nivel .01, ¿sugiere la información que las condiciones del equipo y la política de precios son independientes entre sí? Las cantidades observadas y esperadas estimadas se dan en la tabla 14.10.

Tabla 14.10 Cantidades observadas y esperadas estimadas para el ejemplo 14.14

		Política de precios observada				Política de precios esperada			
		Agresiva	Neutral	No agresiva	$n_{i\cdot}$				
Condición	Abajo de estándar	24	15	17	56	17.02	22.10	16.89	56
	Estándar	52	73	80	205	62.29	80.88	61.83	205
	Moderna	58	86	36	180	54.69	71.02	54.29	180
	$n_{\cdot j}$	134	174	133	441	134	174	133	441

Por tanto

$$\chi^2 = \frac{(24 - 17.02)^2}{17.02} + \cdots + \frac{(36 - 54.29)^2}{54.29} = 22.47$$

y como $\chi^2_{.01,4} = 13.277$, la hipótesis de independencia se rechaza.

La conclusión es que el conocimiento de la política de precios de una gasolinera da información acerca de las condiciones del equipo de la gasolinera. En particular, parece que es más probable que las gasolineras con agresiva política de precios tengan equipo abajo del estándar que las que tienen política neutral o no agresiva. ■

Los modelos y métodos para analizar datos, en los que cada individuo es clasificado con respecto a tres o más factores (tablas de contingencia multidimensionales), se estudian en varias de las referencias de este capítulo.

EJERCICIOS Sección 14.3 (24–36)

24. La tabla bidireccional siguiente se construyó usando datos del artículo "Television Viewing and Physical Fitness in Adults" (*Research Quarterly for Exercise and Sport*, 1990: 315–320). El autor esperaba determinar si el tiempo que la gente pasaba viendo televisión estaba asociado con las condiciones físicas cardiovasculares. A los sujetos se les preguntó sobre sus hábitos de ver televisión y se clasificaron como físicamente en buenas condiciones si entraban en la categoría de excelente o muy bueno en un examen de caminata. Aquí se incluye una salida de Minitab de un análisis de chi cuadrada. Los cuatro grupos que ven televisión corresponden a diferentes cantidades de tiempo por día dedicados a ver TV (0, 1–2, 3–4, o 5 o más horas). Los 168 individuos representados en la primera columna fueron los que se evaluaron en buenas condiciones físicas. Las cantidades esperadas aparecen abajo de las cantidades observadas, y Minitab exhibe la contribución a χ^2 desde cada celda. Exprese y pruebe las hipótesis apropiadas usando $\alpha = .05$.

```
              1            2       Total
1            35          147         182
          25.48       156.52
2           101          629         730
         102.20       627.80
3            28          222         250
          35.00       215.00
4             4           34          38
           5.32        32.68
Total       168         1032        1200

Chi cuadrada = 3.557 + 0.579 +
               0.014 + 0.002 +
               1.400 + 0.228 +
               0.328 + 0.053 = 6.161

gl = 3
```

25. La información siguiente se refiere a marcas en hojas halladas en muestras de trébol blanco seleccionadas de regiones de pastos largos y pastos cortos. ("The Biology of the Leaf Mark Polymorphism in *Trifolium repens* L.", *Heredity*, 1976: 306–325). Use una prueba χ^2 para determinar si las proporciones verdaderas de marcas diferentes son idénticas para los dos tipos de regiones.

	Tipo de marca					Tamaño muestral
	L	LL	Y + YL	O	Otras	
Regiones de pasto largo	409	11	22	7	277	726
Regiones de pasto corto	512	4	14	11	220	761

26. La siguiente información resultó de un experimento para estudiar los efectos del corte de hojas en la capacidad de la fruta de cierto tipo para madurar ("Fruit Set, Herbivory, Fruit Reproduction, and the Fruiting Strategy of *Catalpa speciosa*", *Ecology*, 1980: 57–64):

Tratamiento	Número de frutas maduras	Número de frutas abortadas
Control	141	206
Dos hojas removidas	28	69
Cuatro hojas removidas	25	73
Seis hojas removidas	24	78
Ocho hojas removidas	20	82

¿La información sugiere que la probabilidad de que madure una fruta es afectada por el número de hojas removidas? Exprese y pruebe las hipótesis apropiadas al nivel .01.

27. El artículo "Human Lateralization from Head to Foot: Sex-Related Factors" (*Science*, 1978: 1291–1292) informa que para una muestra de hombres diestros y una muestra de mujeres diestras, el número de individuos cuyos pies eran del mismo tamaño tenían el pie izquierdo más grande que el pie derecho (una diferencia de medio punto en el calzado o más), o tenían el pie derecho más grande que el izquierdo.

	I > D	I = D	I < D	Tamaño muestral
Hombres	2	10	28	40
Mujeres	55	18	14	87

¿La información indica que el género tiene un fuerte efecto en el desarrollo de asimetría en los pies? Exprese las hipótesis nula y alternativa apropiadas, calcule el valor de χ^2, y obtenga información acerca del valor P.

28. Una muestra aleatoria de 175 estudiantes de Cal Poly State University fue seleccionada y tanto el proveedor de servicios de correo electrónico como el proveedor de telefonía celular se determinaron para cada uno, lo que resulta en los datos adjuntos. Establezca y pruebe las hipótesis adecuadas utilizando el método de valor P.

		Provedor de telefonía celular		
		ATT	Verizon	Otro
Proveedor de email	**gmail**	28	17	7
	Yahoo	31	26	10
	Otro	26	19	11

29. Los datos que acompañan el grado de espiritualidad para las muestras de los científicos naturales y sociales en la investigación en las universidades, así como una muestra de no-académicos con títulos de posgrado apareció en el artículo "Conflict Between Religion and Science Among Academic Scientists" (*J. for the Scientific Study of Religion*, 2009: 276–292).

	Grado de espiritualidad			
	Alto	**Moderado**	**Ligero**	**Ninguno**
C.N.	56	162	198	211
C.S.	56	223	243	239
G.A.	109	164	74	28

a. ¿Hay evidencia sustancial para la conclusión de que los tres tipos de individuos no son homogéneos con respecto a su grado de espiritualidad? Establezca y pruebe las hipótesis adecuadas.

b. Teniendo en cuenta sólo a los científicos naturales y científicos sociales, ¿hay evidencia de la falta de homogeneidad? Base su conclusión en un valor *P*.

30. Se consideran tres configuraciones diferentes de diseño para un componente particular. Hay cuatro posibles modos de falla para el componente. Un ingeniero obtuvo la información siguiente sobre el número de fallas en cada modo, para cada una de las tres configuraciones. ¿La configuración parece tener efecto sobre el tipo de falla?

		Modo de falla			
		1	**2**	**3**	**4**
Configuración	**1**	20	44	17	9
	2	4	17	7	12
	3	10	31	14	5

31. Una muestra al azar de los fumadores se obtuvo, y cada individuo se clasificó con respecto al género y con respecto a la edad en que él/ella empezó a fumar. Los datos en la tabla adjunta son consistentes con los resultados reportados en el resumen del artículo "Cigarette Tar Yields in Relation to Mortality in the Cancer Prevention Study II Prospective Cohort" (*British Med. J.*, 2004: 72–79).

		Género	
		Masculino	**Femenino**
Edad	**<16**	25	10
	16 – 17	24	32
	18 – 20	28	17
	>20	19	34

a. Calcule la proporción de hombres en cada categoría de edad, y luego haga lo mismo para las mujeres. Con base en estas proporciones, ¿parece que podría haber una asociación entre el género y la edad a la que un individuo fuma por primera vez?

b. Lleve a cabo una prueba de hipótesis para decidir si existe una asociación entre ambos factores.

32. Cada uno de los estudiantes de una muestra aleatoria de estudiantes, de preparatoria y universidad, se clasificó en cruz con respecto a puntos de vista políticos y consumo de mariguana; los datos resultantes se presentan en la tabla bidireccional siguiente ("Attitudes About Marijuana and Political Views", *Psychological Reports*, 1973: 1051–1054). ¿La información apoya la hipótesis de que los puntos de vista políticos y el nivel de consumo de mariguana son independientes dentro de la población? Pruebe las hipótesis apropiadas usando el nivel de significancia .01.

		Nivel de consumo		
		Nunca	**Rara vez**	**Con frecuencia**
Puntos de vista políticos	**Liberal**	479	173	119
	Conservador	214	47	15
	Otro	172	45	85

33. Demuestre que el estadístico chi cuadrada para la prueba de independencia se puede escribir en la forma

$$\chi^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\frac{N_{ij}^2}{\hat{E}_{ij}}\right) - n$$

¿Por qué esta fórmula es más eficiente computacionalmente que la fórmula de definición para χ^2?

34. Suponga que, en el ejercicio 32, cada uno de los estudiantes ha sido clasificado con respecto a simpatía política, consumo de mariguana y preferencia religiosa, con respecto a las categorías protestante, católica y otras. La información podría exhibirse en tres tablas bidireccionales diferentes, una correspondiente a cada categoría del tercer factor. Con $p_{ijk} = P$(categoría política i, categoría de mariguana j, y categoría religiosa k), la hipótesis nula de independencia de los tres factores expresa que $p_{ijk} = p_{i..}p_{.j.}p_{..k}$. Denote con n_{ijk} la frecuencia observada en celda (i, j, k). Demuestre la forma de estimar las cantidades esperadas de celda suponiendo que H_0 es verdadera ($\hat{e}_{ijk} = n\hat{p}_{ijk}$, de modo que las $\hat{p}_{ijk}$ deban determinarse). A continuación utilice la regla empírica para determinar el número de grados de libertad para el estadístico chi cuadrada.

35. Suponga que en un estado particular formado por cuatro regiones distintas, una muestra aleatoria de n_k votantes se obtiene de la k-ésima región para $k = 1, 2, 3, 4$. Cada votante se clasifica entonces según cuál candidato (1, 2 o 3) prefiere y según el registro del votante (1 = demócrata, 2 = republicano, 3 = independiente). Denote con p_{ijk} la proporción de votantes de la región k que pertenecen a la categoría de candidato i y a la categoría de registro j. La hipótesis nula de regiones homogéneas es H_0: $p_{ij1} = p_{ij2} = p_{ij3} = p_{ij4}$ para toda i, j (es decir, la propor-

ción dentro de la cual cada combinación de candidato/registro es igual para las cuatro regiones). Suponiendo que H_0 es verdadera, determine $\hat{p}_{ijk}$ y $\hat{e}_{ijk}$ como funciones de las n_{ijk} observadas, y use la regla empírica para obtener el número de grados de libertad para la prueba chi cuadrada.

36. Considere la siguiente tabla de 2×3 que presenta las proporciones muestrales que cayeron en las diversas combinaciones de categorías (por ejemplo, 13% de las de la muestra estuvieron en la primera categoría de ambos factores).

	1	2	3
1	.13	.19	.28
2	.07	.11	.22

a. Suponga que la muestra estuvo formada de $n = 100$ personas. Utilice la prueba chi cuadrada para independencia con nivel de significancia de .10.

b. Repita el inciso (a) suponiendo que el tamaño muestral fue de $n = 1000$.

c. ¿Cuál es el tamaño muestral n más pequeño para el que estas proporciones observadas darían por resultado el rechazo de la hipótesis de independencia?

EJERCICIOS SUPLEMENTARIOS (37–49)

37. El artículo "Birth Order and Political Success" (*Psych. Reports*, 1971: 1239-1242) publica que entre 31 candidatos a puestos de elección popular seleccionados al azar, que provenían de familias con cuatro hijos, 12 eran primogénitos, 11 fueron intermedios y 8 fueron últimos. Utilice esta información para probar la hipótesis nula de que es igualmente probable que un candidato político, proveniente de una de estas familias, esté en cualquiera de las cuatro posiciones ordinales.

38. ¿La fase de la Luna tiene alguna incidencia en la tasa de natalidad? Cada uno de 222,784 nacimientos que se produjeron durante un periodo que abarca 24 ciclos lunares completos se clasificó de acuerdo a la fase lunar. Los siguientes datos son consistentes con las cantidades de resumen que aparecieron en el artículo "The Effect of the Lunar Cycle on Frequency of Births and Birth Complications" (*Amer. J. of Obstetrics and Gynecology*, 2005: 1462–1464).

Fase lunar	No. de días en la fase	No. de nacimientos
Luna nueva	24	7680
Luna nueva visible	152	48,442
Cuarto creciente	24	7579
Luna gibosa creciente	149	47,814
Luna llena	24	7711
Luna gibosa menguante	150	47,595
Cuarto menguante	24	7733
Luna menguante	152	48,230

Establezca y pruebe las hipótesis adecuadas para responder a la pregunta planteada al principio de este ejercicio.

39. Las aptitudes de entrenadores en atletismo en jefe y ayudantes en atletismo, hombres y mujeres, se compararon en el artículo "Sex Bias and the Validity of Believed Differences Between Male and Female Interscholastic Athletic Coaches" (*Research Quarterly for Exercise and Sport*, 1990: 259–267). Cada una de las personas en muestras aleatorias de 2225 entrenadores y 1141 entrenadoras, se clasificó según el número de años de experiencia como entrenador para obtener la siguiente tabla bidireccional. ¿Hay suficiente evidencia para concluir que las proporciones que caen en las categorías de experiencia son diferentes para hombres y mujeres? Use $\alpha = .01$.

	Años de experiencia				
Género	**1–3**	**4–6**	**7–9**	**10–12**	**13+**
Masculino	202	369	482	361	811
Femenino	230	251	238	164	258

40. Los autores del artículo "Predicting Professional Sports Game Outcomes from Intermediate Game Scores" (*Chance*, 1992: 18–22) usaron una prueba chi cuadrada, para determinar si había algún mérito en la idea de que los juegos de baloncesto no se deciden hasta el último cuarto, mientras que los juegos de beisbol acaban hacia la séptima entrada. También consideraron futbol y hockey. Se recolectó información de 189 juegos de baloncesto, 92 juegos de beisbol, 80 juegos de hockey y 93 juegos de futbol. Los juegos analizados se muestrearon al azar de todos los encuentros realizados durante la temporada de 1990 para beisbol y futbol, y para la temporada de 1990–1991 de baloncesto y hockey. Para cada juego, se determinó el líder de final de juego y luego se observó si en realidad terminó ganando el partido. La información resultante se resume en la tabla siguiente.

Deporte	Gana líder de final de juego	Pierde líder de final de juego
Baloncesto	150	39
Beisbol	86	6
Hockey	65	15
Futbol	72	21

Los autores expresan que "*líder de final de juego* se define como el equipo que va adelante después del tercer cuarto en baloncesto y futbol, dos tiempos en hockey y siete entradas en beisbol. El valor de chi cuadrada en tres grados de libertad es 10.52 ($P < .015$)".

a. Exprese las hipótesis relevantes y llegue a una conclusión usando $\alpha = .05$.

b. ¿Piensa el lector que su conclusión en el inciso (a) puede atribuirse a que un solo deporte es una anomalía?

41. La siguiente tabla bidireccional de frecuencias aparece en el artículo "Marijuana Use in College" (*Youth and Society*, 1979: 323–334). Cada uno de los 445 estudiantes universitarios fue clasificado según su frecuencia de consumo de mariguana y si los padres consumían alcohol y drogas psicoactivas. ¿La información sugiere que la adicción de padres y de estudiantes son independientes en la población de la que se extrajo la muestra? Utilice el método del valor *P* para llegar a una conclusión.

		Nivel estándar de consumo de mariguana		
		Nunca	**Ocasional**	**Regular**
Padres consumen alcohol y drogas	**Ninguno**	141	54	40
	Uno	68	44	51
	Ambos	17	11	19

42. Se ha puesto mucha atención en la incidencia de las conmociones cerebrales entre los atletas. Las diferentes muestras de jugadores de futbol, los atletas no de futbol, y los no deportistas fueron seleccionados. La tabla adjunta a continuación es el resultado de la determinación del número de conmociones cerebrales de cada individuo reportadas en un cuestionario de historia médica ("No Evidence of Impaired Neurocognitive Performance in Collegiate Soccer Players", *Amer. J. of Sports Med.*, 2002: 157–162).

	No. de conmociones			
	0	**1**	**2**	**≥ 3**
Futbol	45	25	11	10
Atletas NF	68	15	8	5
No atletas	45	5	3	0

¿La distribución de los números de las conmociones cerebrales parece ser diferente para los tres tipos de personas? Lleve a cabo una prueba de hipótesis utilizando el método de valor *P*.

43. En un estudio para investigar el límite al que las personas perciben olores industriales en cierta región ("Annoyance and Health Reactions to Odor from Refineries and Other Industries in Carson, California", *Environmental Research*, 1978: 119–132), se obtuvo una muestra de individuos de cada una de las tres regiones cercanas a instalaciones industriales. A cada uno se le preguntó si percibía olores (1) todos los días, (2) al menos una vez a la semana, (3) al menos una vez al mes, (4) menos de una vez al mes, o (5) nada en absoluto; el resultado es la información de la salida SPSS en la parte superior de la siguiente página. Exprese y pruebe las hipótesis apropiadas.

44. Numerosos compradores han expresado su descontento porque las tiendas de abarrotes han dejado de poner precios en artículos individuales de la tienda. El artículo "The Impact of Item Price Removal on Grocery Shopping Behavior" (*J. of Marketing*, 1980: 73–93) publica un estudio en el que cada comprador de una muestra se clasificó por edad y por si sentía la necesidad que se pusieran precios. Con base en la información siguiente, ¿la necesidad de que se pusieran precios en artículos es independiente de la edad?

	Edad				
	<30	**30–39**	**40–49**	**50–59**	**≥60**
Número en muestra	150	141	82	63	49
Número que desea que se pongan precios	127	118	77	61	41

45. Denote con p_1 la proporción de éxitos en una población particular. El valor del estadístico de prueba del capítulo 8 para probar H_0: $p_1 = p_{10}$ era $z = (\hat{p}_1 - p_{10})/\sqrt{p_{10}p_{20}/n}$, donde $p_{20} = 1 - p_{10}$. Demuestre que para el caso $k = 2$, el valor del estadístico de prueba chi cuadrada de la sección 14.1 satisface $\chi^2 = z^2$. [*Sugerencia:* primero demuestre que $(n_1 - np_{10})^2 = (n_2 - np_{20})^2$.]

46. El torneo de baloncesto de la NCAA empieza con 64 equipos que se reparten en cuatro torneos regionales, cada uno con 16 equipos. Los 16 equipos de cada región se clasifican entonces (sembrados) de 1 a 16. Durante el periodo de 12 años de 1991 a 2002, el equipo clasificado en primer lugar ganó 22 veces su torneo regional, el equipo clasificado en segundo lugar ganó 10 veces, el equipo clasificado en tercer lugar ganó 5 veces, y los 11 torneos regionales restantes fueron ganados por equipos clasificados más abajo del tercer lugar. Denote con P_{ij} la probabilidad de que el equipo clasificado *i* en su región sea ganador en su juego contra el equipo clasificado *j*. Una vez disponibles las P_{ij}, es posible calcular la probabilidad de que cualquier sembrado particular gane su torneo regional (un cálculo complicado porque el número de resultados del espacio muestral es muy grande). El artículo "Probability Models for the NCAA Regional Basketball Tournaments" (*American Statistician*, 1991: 35–38) propuso diversos modelos diferentes para las P_{ij}.

a. Un modelo postuló $P_{ij} = .5 - \lambda(i - j)$ con $\lambda = 1/32$ (de donde $P_{16,1} = \lambda, P_{16,2} = 2\lambda$, etc.). Con base en esto, P(sembrado #1 gana) = .27477, P(sembrado #2 gana) = .20834, y P(sembrado #3 gana) = .15429. ¿Este modelo parece dar un buen ajuste a la información?

SPSS resultado del ejercicio 43
Tabulación cruzada: ÁREA POR CATEGORÍA

CATEGORÍA ÁREA	Conteo Exp Renglón Columna	Val Pct Pct	1.00	2.00	3.00	4.00	5.00	Renglón Total
1.00			20	28	23	14	12	97
			12.7	24.7	18.0	16.0	25.7	33.3%
			20.6%	28.9%	23.7%	14.4%	12.4%	
			52.6%	37.8%	42.6%	29.2%	15.6%	
2.00			14	34	21	14	12	95
			12.4	24.2	17.6	15.7	25.1	32.6%
			14.7%	35.8%	22.1%	14.7%	12.6%	
			36.8%	45.9%	38.9%	29.2%	15.6%	
3.00			4	12	10	20	53	99
			12.9	25.2	18.4	16.3	26.2	34.0%
			4.0%	12.1%	10.1%	20.2%	53.5%	
			10.5%	16.2%	18.5%	41.7%	68.8%	
Columna Total			38	74	54	48	77	291
			13.1%	25.4%	18.6%	16.5%	26.5%	100.0%

Chi-Cuadrada	G.L.	Significancia	Min FE	Celdas con FE < 5
70.64156	8	.0000	12.405	Ninguna

b. Un modelo más refinado tiene probabilidades de juego $P_{ij} = .5 + .2813625\,(z_i - z_j)$, donde las z son medidas de resistencias relativas a percentiles normales estándar [los percentiles para equipos sembrados ganadores están más cerca entre sí que los equipos sembrados más abajo, y .2813625 asegura que el rango de probabilidades es igual para el modelo del inciso (a)]. Las probabilidades resultantes de que los sembrados 1, 2 o 3 ganen sus torneos regionales son .45883, .18813, y .11032, respectivamente. Evalúe el ajuste de este modelo.

47. ¿Se ha preguntado alguna vez si los jugadores de futbol soccer sufren de efectos negativos cuando "cabecean" el balón? Los autores del artículo "No Evidence of Impaired Neurocognitive Performance in Collegiate Soccer Players" (*Amer. J. of Sports Medicine*, 2002: 157–162) investigaron este problema desde varias perspectivas.

a. El artículo informó que 45 de los 91 jugadores de futbol soccer de su muestra sufrieron al menos una concusión, 28 de los 96 atletas no futbolistas habían sufrido al menos una concusión, y sólo 8 de 53 controles de estudiantes habían sufrido al menos una concusión. Analice esta información y saque conclusiones apropiadas.

b. Para los futbolistas, el coeficiente de correlación muestral de los valores de x = exposición al futbol (número total de temporadas de competencia jugadas antes de inscribirse en el estudio) y y = calificación en un examen de recordatorio de memoria inmediata, fue $r = -.220$. Interprete este resultado.

c. A continuación se encuentra un resumen de información sobre calificaciones de un examen oral controlado de asociación de palabras para atletas futbolistas y no futbolistas:

$n_1 = 26,\ \bar{x}_1 = 37.50,\ s_1 = 9.13$
$n_2 = 56,\ \bar{x}_2 = 39.63,\ s_2 = 10.19$

Analice esta información y saque conclusiones apropiadas.

d. Considerando el número de concusiones previas en jugadores no futbolistas, los valores de la media $\pm$ desviación estándar para los tres grupos fueron $.30 \pm .67$, $.49 \pm .87$, y $.19 \pm .48$. Analice esta información y saque conclusiones apropiadas.

48. ¿Los dígitos sucesivos de la expansión decimal de π se comportan como si fueran seleccionados de una tabla de números aleatorios (o de un generador de números aleatorios de una computadora)?

a. Denote por p_0 la proporción a largo plazo de dígitos de la expansión que sean iguales a 0, y defina $p_1, \ldots, p_9$ análogamente. ¿Qué hipótesis acerca de estas proporciones deben probarse, y cuál es el grado de libertad para la prueba de chi cuadrada?

b. H_0 del inciso (a) no sería rechazada para la sucesión no aleatoria 012…901…901…. Considere grupos de dos dígitos que no se traslapen, y denote con p_{ij} la proporción a largo plazo de grupos para los cuales el primer dígito es i y el segundo dígito es j. ¿Qué hipótesis acerca de estas propor-

ciones deben probarse, y cuál es el grado de libertad para la prueba chi cuadrada?

c. Considere grupos de 5 dígitos que no se traslapan. ¿Podría una prueba chi cuadrada de hipótesis apropiadas acerca de las p_{ijklm} estar basada en los primeros 100,000 dígitos? Explique.

d. El artículo "Are the Digits of π an Independent and Identically Distributed Sequence?" (*American Statistician*, 2000: 12–16) consideró los primeros 1,254,540 dígitos de π, y publicó los siguientes valores *P* para tamaños grupales de 1, . . . , 5: .572, .078, .529, .691, .298. ¿Qué concluiría el lector?

49. La secuencia de números de Fibonacci se da en diversos contextos científicos. Los dos primeros números de la secuencia son 1, 1. Entonces, cada número sucesivo es la suma de los dos números anteriores: 1, 1, 1 + 1 = 2, 1 + 2 = 3, 2 + 3 = 5, 8, 13, 21, . . . El primer dígito de un número en esta secuencia puede ser 1, 2, . . . , o 9. Las frecuencias de los primeros dígitos para los primeros 85 números en la secuencia son los siguientes: 25 (1's), 16 (2's), 11, 7, 7, 5, 4, 6, 4. ¿La distribución de los primeros dígitos en la secuencia de Fibonacci parece estar en concordancia con la distribución de la ley de Benford que se describe en el ejercicio 21 del capítulo 3? Establezca y pruebe las hipótesis pertinentes.

Bibliografía

Agresti, Alan, *An Introduction to Categorical Data Analysis*, (2a. ed.), Wiley, New York, 2007. Un excelente examen de diversos aspectos de análisis de datos categóricos por uno de los más prominentes investigadores en este campo.

Everitt, B. S., *The Analysis of Contingency Tables* (2a. edición), Halsted Press, Nueva York, 1992. Un estudio compacto pero informativo de métodos para analizar datos categóricos, expuesto con un mínimo de matemáticas.

Mosteller, Frederick, y Richard Rourke, *Sturdy Statistics*, Addison-Wesley, Reading, MA, 1973. Contiene diversos capítulos fáciles de leer sobre los variados usos de la chi cuadrada.

15 Procedimientos de distribución libre

INTRODUCCIÓN

Cuando la población o poblaciones fundamentales no son normales, las pruebas t y F y los intervalos de confianza t de los capítulos 7–13, tendrán en general niveles de significación reales o niveles de confianza que difieren de los niveles nominales (los prescritos por el experimentador mediante la selección, por ejemplo, de $t_{.025}$, $F_{.01}$, etc.) α y $100(1 - \alpha)\%$], aun cuando la diferencia entre niveles reales y nominales puede no ser grande cuando la desviación desde la normalidad no sea demasiado grande. Debido a que los procedimientos t y F requieren la suposición de distribución de normalidad, no son procedimientos "de distribución libre", alternativamente, porque están basados en una familia paramétrica particular de distribuciones (normal), no son procedimientos "no paramétricos".

En este capítulo se describen procedimientos que son válidos [nivel real α o nivel de confianza $100(1 - \alpha)\%$, simultáneamente para muchos tipos diferentes de distribuciones fundamentales. Estos procedimientos se denominan **de distribución libre** o **no paramétricos.** Se presentan procedimientos de prueba para 1 y 2 muestras en las secciones 15.1 y 15.2 respectivamente. En la sección 15.3 se analizan intervalos de confianza sin distribución. La sección 15.4 describe procedimientos ANOVA sin distribución. Todos estos procedimientos son competidores de los procedimientos paramétricos (t y F), descritos en capítulos anteriores, de modo que es importante comparar el funcionamiento de los dos tipos de procedimientos bajo modelos poblacionales normales y no normales. Hablando en términos generales, los procedimientos de distribución libre funcionan casi igualmente bien como sus similares t y F en el "terreno" de la distribución normal, y con frecuencia dan una mejora considerable bajo condiciones no normales.

15.1 La prueba Wilcoxon de rango con signo

Un químico investigador realizó un experimento químico particular un total de diez veces bajo condiciones idénticas, obteniendo los siguientes valores ordenados de temperatura de reacción:

$$-.57 \quad -.19 \quad -.05 \quad .76 \quad 1.30 \quad 2.02 \quad 2.17 \quad 2.46 \quad 2.68 \quad 3.02$$

Por supuesto que la distribución de temperatura de reacción es continua. Suponga que el investigador está dispuesto a suponer que la distribución de temperatura de reacción es simétrica; es decir, hay un punto de simetría tal que la curva de densidad a la izquierda de ese punto es la imagen en espejo de la curva de densidad a la derecha. Este punto de simetría es la mediana de la distribución (y también es el valor medio μ siempre que la media sea finita). La suposición de simetría puede parecer al principio bastante aventurada, pero recuerde que cualquier distribución normal es simétrica, de modo que la simetría es en realidad una suposición más débil que la normalidad.

Ahora hay que probar la hipótesis nula de que la mediana de la distribución de temperatura de reacción es cero; es decir, H_0: $\tilde{\mu} = 0$. Esto equivale a decir que una temperatura de cualquier magnitud particular, por ejemplo 1.50, no es más probable que sea positiva ($+1.50$) de lo que puede ser negativa (-1.50). Una mirada a los datos sugiere que esta hipótesis no es muy sostenible; por ejemplo, la mediana muestral es 1.66, que es mucho mayor que la magnitud de cualquiera de las tres observaciones negativas.

La figura 15.1 presenta dos diferentes funciones de densidad de probabilidad simétricas, una para la que H_0 es verdadera y una para la que H_a es verdadera. Cuando H_0 es verdadera, se espera que las magnitudes de las observaciones negativas de la muestra sean comparables a las magnitudes de las observaciones positivas. No obstante, si H_0 es "excesivamente" no verdadera como en la figura 15.1(b), entonces las observaciones de magnitud absoluta grande tenderán a ser positivas en lugar de negativas.

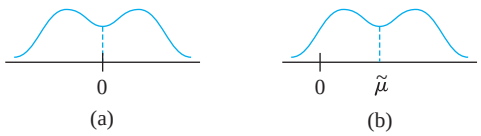


Figura 15.1 Distribuciones para las que (a) $\tilde{\mu} = 0$; (b) $\tilde{\mu} \gg 0$

Para la muestra de diez temperaturas de reacción, por ahora no hay que hacer caso de los signos de observaciones y ordene las magnitudes absolutas de 1 a 10, con la mínima de 1, la segunda más pequeña de 2, y así sucesivamente. A continuación aplique el signo de cada observación al rango correspondiente (de modo que algunos rangos con signo serán negativos, por ejemplo -3, mientras que otros serán positivos, por ejemplo 8). El estadístico de prueba será S_+ = la suma de los rangos con signo positivo.

Magnitud absoluta	.05	.19	.57	.76	1.30	2.02	2.17	2.46	2.68	3.02
Rango	1	2	3	4	5	6	7	8	9	10
Rango con signo	-1	-2	-3	4	5	6	7	8	9	10

$$s_+ = 4 + 5 + 6 + 7 + 8 + 9 + 10 = 49$$

Cuando la mediana de la distribución es mucho mayor a 0, casi todas las observaciones con magnitudes absolutas grandes deberían ser positivas, resultando en rangos de signo positivo y un valor grande de s_+. Por otra parte, si la mediana es 0, las magnitudes de observaciones con signo positivo deberían estar mezcladas con las observaciones de signo negativo, en cuyo caso s_+ no será muy grande. Así, se debería rechazar H_0: $\tilde{\mu} = 0$ en favor de H_a: $\tilde{\mu} > 0$ cuando s_+ es "muy grande" y la región de rechazo debería tener la forma $s_+ \geq c$.

El valor crítico c debería escogerse de modo que la prueba tenga un nivel deseado de significancia (probabilidad de error tipo I) por ejemplo .05 o .01. Para esto es necesario hallar la distribución del estadístico de prueba S_+ cuando la hipótesis nula sea verdadera. Considérese $n = 5$, en cuyo caso hay $2^5 = 32$ formas de aplicar signos a los cinco rangos 1, 2, 3, 4 y 5 (cada rango podría tener un signo $-$ o un signo $+$). El punto clave es que cuando H_0 es verdadera, *cualquier conjunto de cinco rangos con signo tiene la misma probabilidad que cualquier otro conjunto*. Esto es, la observación más pequeña en magnitud absoluta es igualmente probable que sea positiva o negativa, y lo mismo es cierto para la segunda observación más pequeña en magnitud absoluta, y así sucesivamente. Entonces el conjunto $-1, 2, 3, -4, 5$ de rangos con signo es tan probable como el conjunto $1, 2, 3, 4, -5$, y tan probable como cualquiera de las otras 30 posibilidades.

La tabla 15.1 es una lista de 32 posibles secuencias de rango con signo cuando $n = 5$ junto con el valor s_+ para cada secuencia. Esto inmediatamente da la "distribución nula" de S_+ que se ve en la tabla 15.2. Por ejemplo, la tabla 15.1 muestra que tres de las 32 secuencias posibles tienen $s_+ = 8$, de modo que $P(S_+ = 8$ cuando H_0 es verdadera$) = \frac{1}{32} + \frac{1}{32} + \frac{1}{32} = \frac{3}{32}$. Nótese que la distribución nula es simétrica alrededor de 7.5 [en

Tabla 15.1 Posibles secuencias de rango con signo para $n = 5$

Secuencia					s_+	Secuencia					s_+
−1	−2	−3	−4	−5	0	−1	−2	−3	+4	−5	4
+1	−2	−3	−4	−5	1	+1	−2	−3	+4	−5	5
−1	+2	−3	−4	−5	2	−1	+2	−3	+4	−5	6
−1	−2	+3	−4	−5	3	−1	−2	+3	+4	−5	7
+1	+2	−3	−4	−5	3	+1	+2	−3	+4	−5	7
+1	−2	+3	−4	−5	4	+1	−2	+3	+4	−5	8
−1	+2	+3	−4	−5	5	−1	+2	+3	+4	−5	9
+1	+2	+3	−4	−5	6	+1	+2	+3	+4	−5	10
−1	−2	−3	−4	+5	5	−1	−2	−3	+4	+5	9
+1	−2	−3	−4	+5	6	+1	−2	−3	+4	+5	10
−1	+2	−3	−4	+5	7	−1	+2	−3	+4	+5	11
−1	−2	+3	−4	+5	8	−1	−2	+3	+4	+5	12
+1	+2	−3	−4	+5	8	+1	+2	−3	+4	+5	12
+1	−2	+3	−4	+5	9	+1	−2	+3	+4	+5	13
−1	+2	+3	−4	+5	10	−1	+2	+3	+4	+5	14
+1	+2	+3	−4	+5	11	+1	+2	+3	+4	+5	15

Tabla 15.2 Distribución nula de S_+ cuando $n = 5$

s_+	0	1	2	3	4	5	6	7
$p(s_+)$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{2}{32}$	$\frac{2}{32}$	$\frac{3}{32}$	$\frac{3}{32}$	$\frac{3}{32}$
s_+	8	9	10	11	12	13	14	15
$p(s_+)$	$\frac{3}{32}$	$\frac{3}{32}$	$\frac{3}{32}$	$\frac{2}{32}$	$\frac{2}{32}$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{32}$

forma más general, simétricamente distribuida sobre los posibles valores 0, 1, 2, . . . , $n(n + 1)/2$]. Esta simetría es importante para relacionar la región de rechazo de pruebas de cola inferior y de dos colas con la prueba de cola superior.

Para $n = 10$ hay $2^{10} = 1024$ posibles secuencias de rango con signo, de modo que elaborar una lista sería objeto de un gran esfuerzo. Cada secuencia, no obstante, tendría una probabilidad $\frac{1}{1024}$ cuando H_0 es verdadera, de la que la distribución de S_+ cuando H_0 es verdadera se puede obtener fácilmente.

Ahora se puede determinar una región de rechazo para probar H_0: $\tilde{\mu} = 0$ en función de H_a: $\tilde{\mu} > 0$ que tiene un nivel de significancia α adecuadamente pequeño. Considere la región de rechazo $R = \{s_+: s_+ \geq 13\} = \{13, 14, 15\}$. Entonces

$$\begin{aligned}\alpha &= P(\text{rechazar } H_0 \text{ cuando } H_0 \text{ es verdadera})\\ &= P(S_+ = 13, 14, \text{ o } 15 \text{ cuando } H_0 \text{ es verdadera})\\ &= \frac{1}{32} + \frac{1}{32} + \frac{1}{32} = \frac{3}{32}\\ &= .094\end{aligned}$$

de modo que $R = \{13, 14, 15\}$ especifica una prueba con nivel aproximado de .1. La región de rechazo {14, 15}, tiene $\alpha = 2/32 = .063$. Para la muestra $x_1 = .58$, $x_2 = 2.50$, $x_3 = -.21$, $x_4 = 1.23$, $x_5 = .97$, la secuencia de rango con signo es $-1, +2, +3, +4, +5$, de modo que $s_+ = 14$ y al nivel .063 H_0 sería rechazada.

Descripción general de la prueba

Debido a que se supone que la distribución fundamental es simétrica, $\mu = \tilde{\mu}$, de modo que las hipótesis de interés se expresarán en términos de μ más que de $\tilde{\mu}$.*

$$|x_1 - \mu_0|, \ldots, |x_n - \mu_0|$$

SUPOSICIÓN

$X_1, X_2, \ldots, X_n$ es una muestra aleatoria de una distribución de probabilidad continua y simétrica con media (y mediana) μ.

Cuando el valor teorizado de μ es μ_0, las diferencias absolutas deben ordenarse de menor a mayor.

Hipótesis nula: $H_0: \mu = \mu_0$
Valor de estadístico de prueba: $s_+ =$ suma de los rangos asociados con las $(x_i - \mu_0)$ positivas

Hipótesis alternativa	Región de rechazo para prueba de nivel α
$H_a: \mu > \mu_0$	$s_+ \geq c_1$
$H_a: \mu < \mu_0$	$s_+ \leq c_2$ [donde $c_2 = n(n + 1)/2 - c_1$]
$H_a: \mu \neq \mu_0$	$s_+ \geq c$ o $s_+ \leq n(n + 1)/2 - c$

donde los valores críticos c_1 y c obtenidos de la tabla A.13 del apéndice satisfacen $P(S_+ \geq c_1) \approx \alpha$ y $P(S_+ \geq c) \approx \alpha/2$ cuando H_0 es verdadera.

*Si las colas de la distribución son "demasiado gruesas", como fue el caso con la distribución de Cauchy mencionada en el capítulo 6, entonces μ no existirá. En tales casos, la prueba Wilcoxon todavía será válida para pruebas referentes a $\tilde{\mu}$.

Ejemplo 15.1 Un fabricante de planchas eléctricas, deseando probar la precisión del control del termostato en la posición de ajuste de 500°F, da instrucciones a una ingeniera de pruebas para obtener temperaturas reales en esa posición de ajuste para 15 planchas que usan termopar. Las mediciones resultantes son como sigue:

494.6 510.8 487.5 493.2 502.6 485.0 495.9 498.2
501.6 497.3 492.0 504.3 499.2 493.5 505.8

La ingeniera piensa que es razonable suponer que una desviación de temperatura de cualquier magnitud, desde los 500°, es probable que sea tanto positiva como negativa (la suposición de simetría), pero desea protegerse contra una posible situación fuera de normalidad de la distribución real de temperatura, de modo que decide usar la prueba Wilcoxon de rango con signo para ver si la información da una fuerte sugerencia de calibración incorrecta de la plancha.

Las hipótesis son: H_0: $\mu = 500$ en función de H_a: $\mu \neq 500$, donde μ = promedio verdadero de temperatura real en la posición de ajuste de 500°F. Restando 500 de cada x_i se tiene

−5.6 10.8 −12.5 −6.8 2.6 −15.0 −4.1 −1.8 1.6 −2.7
−8.0 4.3 −.8 −6.5 5.8

Los rangos se obtienen al ordenar éstos de menor a mayor cualquiera que sea su signo.

Magnitud absoluta	.8	1.6	1.8	2.6	2.7	4.1	4.3	5.6	5.8	6.5	6.8	8.0	10.8	12.5	15.0
Rango	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Signo	−	+	−	+	−	−	+	−	+	−	−	−	+	−	−

Así, $s_+ = 2 + 4 + 7 + 9 + 13 = 35$. De la tabla A.13 del apéndice, $P(S_+ \geq 95) = P(S_+ \leq 25) = .024$ cuando H_0 es verdadera, de modo que la prueba de dos colas con nivel aproximado de .05 rechaza H_0 cuando $s_+ \geq 95$ o ≤ 25 [la α exacta es 2(.024) = .048]. Como $s_+ = 35$ no está en la región de rechazo, no se puede concluir al nivel .05 que μ sea algo diferente a 500. Incluso al nivel .094 (aproximadamente .1), H_0 no es rechazada, porque $P(S_+ \leq 30) = .047$ implica que valores s_+ entre 30 y 90 no son importantes a ese nivel. El valor P de los datos es entonces mayor a .1. ■

Aun cuando una implicación teórica de la continuidad de la distribución fundamental es que no ocurrirán empates, es frecuente que en la práctica sí ocurran debido a lo discreto de los instrumentos de medida. Si hay diversos valores de información con la misma magnitud absoluta, entonces se les asignaría el promedio de los rangos que recibirían si fueran muy ligeramente distintos entre sí. Así, en el ejemplo 15.1 $x_8 = 498.2$ se cambia a 498.4, entonces dos valores diferentes de $(x_i - 500)$ tendrían magnitud absoluta de 1.6. Los rangos a ser promediados serían 2 y 3, por lo que a cada uno se le asignaría rango 2.5.

Observaciones pareadas

Cuando los datos consistieron en $(X_1, Y_1), \ldots, (X_n, Y_n)$ y las diferencias $D_1 = X_1 - Y_1, \ldots, D_n = X_n - Y_n$ estuvieron normalmente distribuidas, en el capítulo 9 se usó una prueba t pareada para probar hipótesis acerca de la diferencia esperada μ_D. Si no se supone normalidad, las hipótesis acerca de μ_D se pueden probar usando la prueba Wilcoxon de rango con signo en las D_i siempre que la distribución de las diferencias sea continua y simétrica. *Si X_i y Y_i tienen distribuciones continuas que difieren sólo con respecto a sus medias, entonces D_i ten-*

drá una distribución simétrica continua (*no* es necesario que las distribuciones X y Y sean simétricas individualmente). La hipótesis nula es H_0: $\mu_D = \Delta_0$, y el estadístico de prueba S_+ es la suma de los rangos asociados con las $(D_i - \Delta_0)$ positivas.

Ejemplo 15.2 El ayuno intermitente (AI) consiste en episodios repetidos de ayuno a corto plazo. Es de interés potencial, ya que puede proporcionar una herramienta simple para mejorar la sensibilidad a la insulina en individuos con resistencia a ésta (esta última aumenta la probabilidad de diabetes tipo 2 y enfermedades cardiacas). El artículo "Intermittent Fasting Does Not Affect Whole-Body Glucose, Lipid, or Protein Metabolism" (*Amer. J. of Clinical Nutr.*, 2009: 1244–1251) informó sobre un estudio en el que el gasto energético en reposo (kcal/día) determinado en una muestra de $n = 8$ sujetos, que están en un régimen de AI y en una dieta estándar. Los autores del artículo amablemente dieron los siguientes datos:

Sujeto	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Régimen AI	1753.7	1604.4	1576.5	1279.7	1754.2	1695.5	1700.1	1717.0
Dieta estándar	1755.0	1691.1	1697.1	1477.7	1785.2	1669.7	1901.3	1735.3
Diferencia	−1.3	−86.7	−120.6	−198.0	−31.0	25.8	−201.2	−18.3
Rango con signo	−1	−5	−6	−7	−4	3	−8	−2

El artículo utilizó en las diferencias la prueba de rangos con signo de Wilcoxon, para decidir si existe alguna diferencia entre el REE promedio real para el régimen AI y la dieta estándar. Las hipótesis relevantes son

$$H_0: \mu_D = 0 \text{ contra } H_a: \mu_D \neq 0$$

El valor estadístico de prueba es claramente $s_+ = 3$ (sólo ese rango con signo es positivo). Para una prueba de dos colas, el valor P es $2P(S_+ \leq 3$ cuando H_0 es verdadera). En el caso $n = 8$, hay $2^8 = 256$ posibles conjuntos de rangos con signo, todos los cuales tienen la misma probabilidad de que la hipótesis nula sea cierta. Los conjuntos de rangos con signo que resultan de los valores del estadístico de prueba son menores o más pequeños que el valor 3 que vino de los datos como sigue (sólo se muestran rangos con signo positivos):

rangos con signo no positivo ($s_+ = 0$); 1 ($s_+ = 1$); 2 ($s_+ = 2$); 1, 2 ($s_+ = 3$); 3 ($s_+ = 3$)

Así que el valor P es 2(5/256) = .039. Por lo tanto, la hipótesis nula sería rechazada al nivel de significancia .05, pero no al de .01. El artículo informaba sólo que el valor $P < .05$.

Aquí está la salida del paquete de software R:

```
Prueba de rangos con signos de Wilcoxon
datos: y
V = 3, valor p = 0.03906
hipótesis alternativa: la posición verdadera no es igual a 0
Prueba de rangos con signo con corrección continua
datos: y
V = 3, valor p = 0.04232
hipótesis alternativa: la posición verdadera no es igual a 0
```

Este último valor P de .042, que también reporta Minitab, se basa en la aproximación normal descrita en la subsección siguiente, junto con una corrección de continuidad. ■

Una aproximación a muestra grande

La tabla A.13 del apéndice da valores críticos para pruebas α sólo cuando $n \leq 20$. Para $n > 20$, se puede demostrar que S_+ tiene aproximadamente una distribución normal con

$$\mu_{S_+} = \frac{n(n+1)}{4} \qquad \sigma^2_{S_+} = \frac{n(n+1)(2n+1)}{24}$$

cuando H_0 es verdadera.

La media y varianza resultan de observar que cuando H_0 es verdadera (la distribución simétrica está centrada en μ_0), entonces es igualmente probable que el rango i reciba un signo $+$ que un signo $-$. Así,

$$S_+ = W_1 + W_2 + W_3 + \cdots + W_n$$

donde

$$W_1 = \begin{cases} 1 & \text{con probabilidad .5} \\ 0 & \text{con probabilidad .5} \end{cases} \quad \cdots \quad W_n = \begin{cases} n & \text{con probabilidad .5} \\ 0 & \text{con probabilidad .5} \end{cases}$$

($W_i = 0$ es equivalente a que el rango i sea asociado con un $-$, de modo que i no contribuye a S_+.)

S_+ es entonces una suma de variables aleatorias, y cuando H_0 es verdadera, se puede demostrar que estas W_i son independientes. La aplicación de las reglas de valor esperado y varianza da la media y varianza de S_+. Como las W_i no están distribuidas de manera idéntica, nuestra versión del teorema de límite central no se puede aplicar, pero hay una versión más general del teorema que se puede usar para justificar la conclusión de normalidad. Al poner estos resultados juntos tenemos la siguiente prueba estadística de muestra grande.

$$Z = \frac{S_+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}} \tag{15.1}$$

Para las tres alternativas estándar, los valores críticos para pruebas de nivel α son los valores normales estándares usuales z_α, $-z_\alpha$ y $\pm z_{\alpha/2}$.

Ejemplo 15.3 Un tipo particular de viga de acero se ha diseñado para tener una resistencia a la compresión (lb/pulg^2) de al menos 50,000. Para cada viga de una muestra de 25 vigas, la resistencia a la compresión se determinó y aparece en la tabla 15.3. Suponiendo que la resistencia real a la compresión está distribuida simétricamente alrededor del verdadero valor prome-

Tabla 15.3 Datos para el ejemplo 15.3

$x_i - 50{,}000$	Rango con signo	$x_i - 50{,}000$	Rango con signo	$x_i - 50{,}000$	Rango con signo
−10	−1	−99	−10	165	+18
−27	−2	113	+11	−178	−19
36	+3	−127	−12	−183	−20
−55	−4	−129	−13	−192	−21
73	+5	136	+14	−199	−22
−77	−6	−150	−15	−212	−23
−81	−7	−155	−16	−217	−24
90	+8	−159	−17	−229	−25
−95	−9				

dio, utilice la prueba Wilcoxon para determinar si la verdadera resistencia promedio a la compresión es menor que el valor especificado. Esto es, pruebe H_0: $\mu = 50{,}000$ en función de H_a: $\mu < 50{,}000$ (a favor de la afirmación de que el promedio de resistencia a la compresión es al menos 50,000).

La suma de los rangos con signo positivo es $3 + 5 + 8 + 11 + 14 + 18 = 59$, $n(n + 1)/4 = 162.5$ y $n(n + 1)(2n + 1)/24 = 1381.25$, de modo que

$$z = \frac{59 - 162.5}{\sqrt{1381.25}} = -2.78$$

La prueba de nivel .01 de cola inferior rechaza H_0 si $z \leq -2.33$. Como $-2.78 \leq -2.33$, H_0 es rechazada a favor de la conclusión de que el promedio verdadero de resistencia a la compresión es menor a 50,000. ■

Cuando hay empates en las magnitudes absolutas, de modo que deban usarse rangos promedio, todavía es correcto estandarizar S_+ al restar $n(n + 1)/4$, pero debe usarse la siguiente fórmula corregida para varianza:

$$\sigma^2_{S_+} = \frac{1}{24} n(n + 1)(2n + 1) - \frac{1}{48} \sum(\tau_i - 1)(\tau_i)(\tau_i + 1) \qquad (15.2)$$

donde τ_i es el número de empates en el i-ésimo conjunto de valores empatados y la suma sea mayor que todos los conjuntos de valores empatados. Si, por ejemplo, $n = 10$ y los rangos con signo son 1, 2, -4, -4, 4, 6, 7, 8.5, 8.5, y 10, entonces hay dos conjuntos empatados con $\tau_1 = 3$ y $\tau_2 = 2$, de modo que la suma es $(2)(3)(4) + (1)(2)(3) = 30$ y $\sigma^2_{S_+} = 96.25 - 30/48 = 95.62$. El denominador de (15.1) debe ser sustituido por la raíz cuadrada en (15.2), aun cuando, como muestra este ejemplo, la corrección suele ser insignificante.

Eficiencia de la prueba Wilcoxon de los rangos con signo

Cuando la distribución fundamental que se muestrea es normal, se puede usar la prueba t o la prueba de rangos con signo para probar una hipótesis acerca de μ. La prueba t es la mejor en esta situación porque entre todos los niveles de pruebas α es la que tiene β mínima. Puesto que en general se conviene que hay numerosas situaciones experimentales en las que la normalidad se puede suponer razonablemente, así como algunas en las que esto no debería ser, hay dos preguntas que deben abordarse en un intento por comparar las dos pruebas:

1. Cuando la distribución fundamental es normal (el "terreno" de la prueba t), ¿cuánto se pierde al usar la prueba de rangos con signo?
2. Cuando la distribución fundamental no es normal, ¿puede lograrse una mejoría importante al usar la prueba de rangos con signo?

Si la prueba Wilcoxon no se afecta mucho con respecto a la prueba t en el "terreno" de esta última, y funciona considerablemente mejor que la prueba t para un gran número de otras distribuciones, entonces habrá una fuerte inclinación para usar la prueba Wilcoxon.

Por desgracia, no hay respuestas sencillas a estas preguntas. Al reflexionar, no es de sorprenderse que la prueba t pueda funcionar mal cuando la distribución fundamental tenga "colas gruesas" (es decir, cuando valores observados que se encuentren lejos de μ sean relativamente más probables de lo que son cuando la distribución es normal). Esto es porque el comportamiento de la prueba t depende de la media muestral, que puede ser muy inestable en presencia de colas gruesas. La dificultad en producir respuestas a las dos preguntas es que β para la prueba Wilcoxon es muy difícil de obtener y estudiar para *cualquier* distribución fundamental, y lo mismo se puede decir para la prueba t cuando la distribución no sea normal. Incluso si β pudiera obtenerse con facilidad, cualquier medida de eficiencia dependería claramente de qué distribución fundamental se postuló. Expertos

en estadística han propuesto varias medidas diferentes de eficiencia; una que muchos de éstos consideran creíble se denomina **eficiencia asintótica relativa** (ARE, por sus siglas en inglés). La ARE de una prueba con respecto a otra es en esencia la razón limitadora de tamaños muestrales para obtener idénticas probabilidades de error para las dos pruebas. Entonces, si la ARE de una prueba con respecto a una segunda es .5, entonces cuando los tamaños muestrales son grandes, se requerirá un tamaño muestral el doble de grande de la primera prueba para que funcione tan bien como la segunda prueba. Aun cuando la ARE no caracteriza una operación de prueba para tamaños de muestra pequeños, se puede demostrar que se cumplen los siguientes resultados:

1. Cuando la distribución fundamental es normal, la ARE de la prueba Wilcoxon con respecto a la prueba t es aproximadamente .95.
2. Para cualquier distribución, la ARE será al menos .86 y para muchas distribuciones será mucho mayor a 1.

Se pueden resumir estos resultados diciendo que, en problemas de muestras grandes, la prueba Wilcoxon nunca es mucho menos eficiente que la prueba t y puede ser mucho más eficiente si la distribución fundamental está lejos de ser normal. Aun cuando el problema está lejos de ser resuelto en el caso de tamaños muestrales obtenidos en casi todos los problemas prácticos, estudios hechos han demostrado que la prueba Wilcoxon funciona razonablemente y por ello es una alternativa viable para la prueba t.

EJERCICIOS Sección 15.1 (1–9)

1. Reconsidere la situación descrita en el ejercicio 34 de la sección 8.2, y use la prueba Wilcoxon con $\alpha = .05$ para probar las hipótesis relevantes.

2. Sean otra vez los datos del índice de gasto (%) para una muestra de 20 fondos de inversión mezclados de gran capitalización presentados en el ejercicio 1.53:

```
1.03   1.23   1.10   1.64   1.30   1.27   1.25
 .78   1.05    .64    .94    .86   1.05    .75
 .09   0.79   1.61   1.26    .93    .84
```

Una gráfica de probabilidad normal muestra un patrón claramente no lineal, principalmente debido a la única atípica en cada extremo de los datos. Sin embargo, una gráfica de puntos y un diagrama de caja muestran una cantidad razonable de simetría. Suponiendo una distribución simétrica de la población, ¿los datos proporcionan pruebas convincentes para concluir que la media poblacional del índice de gastos es superior al 1%? Use la prueba de Wilcoxon al nivel de significancia .1. [*Nota*: la tasa de gasto medio de la población de los 825 fondos en realidad es 1.08.]

3. La información siguiente es un subconjunto de la publicada en el artículo "Synovial Fluid pH, Lactate, Oxygen and Carbon Dioxide Partial Pressure in Various Joint Diseases" (*Arthritis and Rheumatism*, 1971: 476–477). Las observaciones son valores de pH del fluido sinovial (que lubrica articulaciones y tendones) tomado de las rodillas de personas que sufren de artritis. Suponiendo que el pH promedio real para personas no artríticas es de 7.39, pruebe al nivel .05 para ver si la información indica una diferencia entre valores del pH promedio para personas artríticas y no artríticas.

7.02 7.35 7.34 7.17 7.28 7.77 7.09
7.22 7.45 6.95 7.40 7.10 7.32 7.14

4. Se seleccionó una muestra aleatoria de 15 mecánicos para automóviles, certificados para trabajar en cierto tipo de auto, y se determinó el tiempo (en minutos) necesario para que cada uno de ellos diagnosticara un problema particular; la información resultante fue la siguiente:

30.6 30.1 15.6 26.7 27.1 25.4 35.0 30.8
31.9 53.2 12.5 23.2 8.8 24.9 30.2

Utilice la prueba Wilcoxon en el nivel de significancia .10 para determinar si la información sugiere que el promedio verdadero del tiempo de diagnóstico es menor a 30 minutos.

5. Un método gravimétrico y uno espectrofotométrico están bajo consideración para determinar el contenido de fosfato de un material particular. Se obtienen 12 muestras del material, cada una se corta a la mitad, y se hace la determinación de cada mitad usando uno de los dos métodos; los datos resultantes son los siguientes:

Muestra	1	2	3	4
Gravimétrico	54.7	58.5	66.8	46.1
Espectrofotométrico	55.0	55.7	62.9	45.5
Muestra	5	6	7	8
Gravimétrico	52.3	74.3	92.5	40.2
Espectrofotométrico	51.1	75.4	89.6	38.4
Muestra	9	10	11	12
Gravimétrico	87.3	74.8	63.2	68.5
Espectrofotométrico	86.8	72.5	62.3	66.0

Use la prueba Wilcoxon para determinar si una técnica da, en promedio, un valor diferente al de la otra técnica para este tipo de material.

6. Reconsidere la situación descrita en el ejercicio 39 de la sección 9.3, y use la prueba Wilcoxon para probar las hipótesis apropiadas.

7. Use la versión de muestra grande de la prueba Wilcoxon en el nivel de significancia .05, en los datos del ejercicio 37 de la sección 9.3, para determinar si la diferencia media verdadera entre concentraciones interiores y exteriores es mayor a .20.

8. Reconsidere los datos del alcohol contenido en el Oporto del ejercicio 14.22. Una gráfica de probabilidad normal arroja algunas dudas sobre el supuesto de normalidad de la población. Sin embargo, una gráfica de puntos muestra una cantidad razonable de simetría y la media, la mediana y la media recortada al 5% son 19.257, 19.200 y 19.209, respectivamente. Use la prueba Wilcoxon al nivel de significancia .01 para decidir si existe evidencia sustancial para concluir que el contenido promedio real excede 18.5.

9. Suponga que se hacen observaciones $X_1, X_2, \ldots, X_n$ en un proceso, en tiempos $1, 2, \ldots, n$. Con base en estos datos, se desea probar

H_0: las X_i constituyen una secuencia independiente y distribuida de manera idéntica

en función de

H_a: X_{i+1} tiende a ser mayor que X_i para $i = 1, \ldots, n$ (una tendencia creciente)

Suponga que las X_i están ordenadas de 1 a n. Entonces, cuando H_a es verdadera, los rangos más grandes tienden a presentarse más tarde en la secuencia, en tanto que si H_0 es verdadera, los rangos grandes y pequeños tienden a mezclarse. Sea R_i el rango de X_i y considere el estadístico de prueba $D = \sum_{i=1}^{n}(R_i - i)^2$. Entonces los valores pequeños de D dan apoyo a H_a (por ejemplo, el valor más pequeño es 0 para $R_1 = 10$, $R_2 = 2, \ldots, R_n = n$), de modo que H_0 debe ser rechazada a favor de H_a si $d \leq c$. Cuando H_0 es verdadera, cualquier secuencia de rangos tiene probabilidad $1/n!$ Use esto con el fin de hallar c para la que tiene un nivel tan cercano a .10 como es posible en el caso $n = 4$. [*Sugerencia:* haga una lista de las secuencias de rango 4!, calcule d para cada una, y luego obtenga la distribución nula de D. Vea el libro de Lehmann (en la bibliografía de este capítulo), p. 290, para más información.]

15.2 Prueba Wilcoxon de suma de rangos

La prueba t de dos muestras se basa en el supuesto de que ambas distribuciones poblacionales son normales. Hay situaciones, sin embargo, en las que un investigador desearía usar una prueba válida incluso si las distribuciones fundamentales son bastante no normales. A continuación se describe esa prueba, llamada **prueba Wilcoxon de suma de rangos.** Un nombre alternativo para el procedimiento es prueba Mann-Whitney, aun cuando el estadístico de la prueba Mann-Whitney se expresa a veces en una forma ligeramente diferente de la prueba Wilcoxon. El procedimiento de la prueba Wilcoxon es de distribución libre porque tendrá el nivel deseado de significancia para una clase muy grande de distribuciones fundamentales.

SUPOSICIONES

$X_1, \ldots, X_m$ y $Y_1, \ldots, Y_n$ son dos muestras aleatorias independientes de distribuciones continuas con medias μ_1 y μ_2, respectivamente. Las distribuciones X y Y tienen la misma forma y dispersión, con la única diferencia posible entre las dos estando en los valores de μ_1 y μ_2.

La hipótesis nula H_0: $\mu_1 - \mu_2 = \Delta_0$ afirma que la distribución X es desplazada por una cantidad Δ_0 a la derecha de la distribución Y.

Desarrollo de la prueba cuando $m = 3, n = 4$

Considere probar primero H_0: $\mu_1 - \mu_2 = 0$. Si μ_1 es en realidad mucho mayor que μ_2, entonces casi todas las x observadas caerán a la derecha de las y observadas. No obstante, si H_0 es verdadera, entonces los valores observados de las dos muestras deben estar entremezclados. El estadístico de prueba dará una cuantificación de cuánta mezcla hay en las dos muestras.

Considere el caso $m = 3$, $n = 4$. Entonces si las tres x observadas estuvieran a la derecha de las cuatro y observadas, esto sería una fuerte evidencia para rechazar H_0 a favor

de H_a: $\mu_1 - \mu_2 \neq 0$; una conclusión semejante es apropiada si las tres x caen debajo de las cuatro y. Suponga que se agrupan las X y las Y en una muestra combinada de tamaño $m + n = 7$ y se ordenan estas observaciones de menor a mayor, con la más pequeña recibiendo el rango 1 y la mayor el rango 7. Si casi todos los rangos más grandes o los rangos más pequeños se asociaran con observaciones X, se empezaría a dudar de H_0. Esto sugiere el estadístico de prueba

$$W = \text{la suma de los rangos de la muestra combinada asociada con observaciones de } X \tag{15.3}$$

Para los valores de m y n bajo consideración, el valor más pequeño posible de W es $w = 1 + 2 + 3 = 6$ (si las tres x son menores que las cuatro y), y el máximo valor posible es $w = 5 + 6 + 7 = 18$ (si las tres x son mayores que las cuatro y).

Como ejemplo, suponga que $x_1 = -3.10, x_2 = 1.67, x_3 = 2.01, y_1 = 5.27, y_2 = 1.89, y_3 = 3.86$, y $y_4 = .19$. Entonces la muestra ordenada agrupada es -3.10, .19, 1.67, 1.89, 2.01, 3.86 y 5.27. Los rangos X para esta muestra son 1 (para -3.10), 3 (para 1.67), y 5 (para 2.01), de modo que el valor calculado de W es $w = 1 + 3 + 5 = 9$.

El procedimiento de prueba basado en el estadístico (15.3) es rechazar H_0 si el valor calculado de w es "demasiado extremo", es decir, $\geq c$ para una prueba de cola superior, $\leq c$ para una prueba de cola inferior, y $\geq c_1$ o $\leq c_2$ para una prueba de dos colas. La(s) constante(s) crítica(s) c (c_1, c_2) deben escogerse de modo que la prueba tenga el nivel deseado de significancia α. Para ver cómo debería hacerse esto, recuerde que cuando H_0 es verdadera, las siete observaciones provienen de la misma población. Esto significa que bajo H_0, cualquier posible terna de rangos asociada con las tres x, por ejemplo (1, 4, 5), (3, 5, 6) o (5, 6, 7) tiene la misma probabilidad que cualquier otra posible terna de rangos. Como hay $\binom{7}{3} = 35$ posibles ternas de rangos, bajo H_0 cada terna de rangos tiene probabilidad $\frac{1}{35}$. De una lista de las 35 ternas de rangos y el valor w asociado con cada una, la distribución de probabilidad de W puede determinarse de inmediato. Por ejemplo, hay cuatro ternas de rangos que tienen valor w de 11, (1, 3, 7), (1, 4, 6), (2, 3, 6) y (2, 4, 5), por lo que $P(W = 11) = \frac{4}{35}$. El resumen de la lista y cálculos aparecen en la tabla 15.4.

Tabla 15.4 Distribución de probabilidad de W ($m = 3, n = 4$) cuando H_0 es verdadera

w	6	7	8	9	10	11	12	13	14	15	16	17	18
$P(W = w)$	$\frac{1}{35}$	$\frac{1}{35}$	$\frac{2}{35}$	$\frac{3}{35}$	$\frac{4}{35}$	$\frac{4}{35}$	$\frac{5}{35}$	$\frac{4}{35}$	$\frac{4}{35}$	$\frac{3}{35}$	$\frac{2}{35}$	$\frac{1}{35}$	$\frac{1}{35}$

La distribución de la tabla 15.4 es simétrica alrededor del valor $w = (6 + 18)/2 = 12$, que es el valor central de la lista ordenada de posibles valores de W. Esto es porque las dos ternas de rangos (r, s, t) (con $r < s < t$) y ($8 - t$, $8 - s$, $8 - r$) tienen valores de w simétricos alrededor de 12, de modo que para cada terna con valor w debajo de 12, hay una terna con valor w arriba de 12 en la misma cantidad.

Si la hipótesis alternativa es H_a: $\mu_1 - \mu_2 > 0$, entonces H_0 debe ser rechazada a favor de H_a para valores W grandes. Si se escoge como la región de rechazo al conjunto de valores W {17, 18}, $\alpha = P$(tipo I de error)$= P$(rechazar H_0 cuando H_0 es verdadera) $= P(W =$ 17 o 18 cuando H_0 es verdadera)$= \frac{1}{35} + \frac{1}{35} = \frac{2}{35} = .057$; la región {17, 18} por tanto especifica una prueba con nivel de significancia de alrededor de .05. Del mismo modo, la región {6, 7}, que es apropiada para H_a: $\mu_1 - \mu_2 < 0$, tiene $\alpha = .057 \approx .05$. La región {6, 7, 17, 18}, que es apropiada para la alternativa de dos lados, tiene $\alpha = \frac{4}{35} = .114$. El valor W para la información dada varios párrafos atrás era $w = 9$, que está más bien cerca

del valor central 12, de modo que H_0 no sería rechazada a ningún nivel α razonable para cualquiera de las tres H_a.

Descripción general de la prueba

La hipótesis nula H_0: $\mu_1 - \mu_2 = \Delta_0$ se maneja restando Δ_0 de cada X_i y usando las $(X_i - \Delta_0)$ como las X_i se usaron previamente. Recordando que para cualquier entero positivo K, la suma de los primeros K enteros es $K(K + 1)/2$, el mínimo valor posible del estadístico W es $m(m + 1)/2$, que se presenta cuando las $(X_i - \Delta_0)$ están todas a la izquierda de la muestra Y. El máximo valor posible de W se presenta cuando las $(X_i - \Delta_0)$ están por completo a la derecha de las Y; en este caso, $W = (n + 1) + \cdots + (m + n) =$ (suma de los primeros $m + n$ enteros) $-$ (suma de los primeros n enteros), que da $m(m + 2n + 1)/2$. Al igual que con el caso especial $m = 3$, $n = 4$, la distribución de W es simétrica alrededor del valor que está a la mitad entre los valores mínimo y máximo; este valor central es $m(m + n + 1)/2$. Debido a esta simetría, las probabilidades que comprenden valores críticos de cola inferior se pueden obtener de los correspondientes valores de cola superior.

Hipótesis nula: H_0: $\mu_1 - \mu_2 = \Delta_0$

Valor estadístico de prueba: $w = \sum_{i=1}^{m} r_i$ donde $r_i =$ rango de $(x_i - \Delta_0)$ en la muestra combinada de las $m + n$ $(x - \Delta_0)$ y las y

Hipótesis alternativa	Región de rechazo
H_a: $\mu_1 - \mu_2 > \Delta_0$	$w \geq c_1$
H_a: $\mu_1 - \mu_2 < \Delta_0$	$w \leq m(m + n + 1) - c_1$
H_a: $\mu_1 - \mu_2 \neq \Delta_0$	$w \geq c$ o $w \leq m(m + n + 1) - c$

donde $P(W \geq c_1$ cuando H_0 es verdadera$) \approx \alpha$, $P(W \geq c$ cuando H_0 es verdadera$) \approx \alpha/2$.

Debido a que W tiene una distribución discreta, no siempre existirá un valor crítico que corresponda exactamente a uno de los niveles usuales de significancia. La tabla A.14 del apéndice da valores críticos de cola superior para probabilidades más cercanas a .05, .025, .01, y .005, de cuyos niveles .05 o .01 se pueden obtener pruebas de una y dos colas. La tabla da información sólo para $m = 3, 4, \ldots, 8$ y $n = m, m + 1, \ldots, 8$ (es decir, $3 \leq m \leq n \leq 8$). Para valores de m y n que excedan de 8, se puede usar una aproximación normal. No obstante, para usar la tabla para m y n pequeñas *las muestras X y Y deben marcarse para que* $m \leq n$.

Ejemplo 15.4 La concentración de fluoruro urinario (partes por millón) se midió para una muestra de ganado que pastaba en una zona previamente expuesta a contaminación de fluoruro, y para una muestra similar que pastaba en una región sin contaminación:

Contaminado	21.3	18.7	23.0	17.1	16.8	20.9	19.7
No contaminado	14.2	18.3	17.2	18.4	20.0		

¿Esta información indica firmemente que el promedio verdadero de concentración de fluoruro, para ganado que pasta en la región contaminada, es mayor que para la región no contaminada? Utilice la prueba Wilcoxon de suma de rangos al nivel $\alpha = .01$.

Los tamaños muestrales en este caso son 7 y 5. Para obtener $m \leq n$, marque las observaciones no contaminadas como las x ($x_1 = 14.2, \ldots, x_5 = 20.0$). De este modo, μ_1

es el promedio verdadero de concentración de fluoruro sin contaminación y μ_2 es el promedio verdadero de concentración con contaminación. La hipótesis alternativa es H_a: $\mu_1 - \mu_2 < 0$ (la contaminación está asociada con un aumento en concentración), de modo que es apropiada una prueba de cola inferior. De la tabla A.14 del apéndice con $m = 5$ y $n = 7$, $P(W \geq 47$ cuando H_0 es verdadera) $\approx .01$. El valor crítico para la prueba de cola inferior es por tanto $m(m + n + 1) - 47 = 5(13) - 47 = 18$; H_0 será rechazada ahora si $w \leq 18$. Sigue la muestra ordenada agrupada; la W calculada es $w = r_1 + r_2 + \cdots + r_5$ (donde r_i es el rango de x_i) $= 1 + 5 + 4 + 6 + 9 = 25$. Como $25 > 18$, H_0 no es rechazada al (aproximadamente) nivel .01.

x	*y*	*y*	*x*	*x*	*x*	*y*	*y*	*x*	*y*	*y*	*y*
14.2	16.8	17.1	17.2	18.3	18.4	18.7	19.7	20.0	20.9	21.3	23.0
1	2	3	4	5	6	7	8	9	10	11	12

Teóricamente, la suposición de continuidad de las dos distribuciones asegura que todas las $(m + n)$ x y y observadas tengan valores diferentes. En la práctica, no obstante, con frecuencia habrá empates en los valores observados. Al igual que con la prueba Wilcoxon de rango con signo, la práctica común al trabajar con empates es asignar, a cada una de las observaciones enlazadas de un conjunto particular de empates, el promedio de los rangos que recibirían si fueran ligeramente diferentes entre sí.

Una aproximación normal para *W*

Cuando m y n exceden de 8, la distribución de W puede ser aproximada por una curva normal apropiada, y esta aproximación se puede usar en lugar de la tabla A.14 del apéndice. Para obtener la aproximación, se necesita μ_W y σ_W^2 cuando H_0 es verdadera. En este caso, el rango R_i de $X_i - \Delta_0$ es igualmente probable que sea cualquiera de los posibles valores $1, 2, 3, \ldots, m + n$ (R_i tiene una distribución uniforme discreta en los primeros $m + n$ enteros positivos), de modo que $\mu_{R_i} = (m + n + 1)/2$. Como $W = \sum R_i$, esto da

$$\mu_W = \mu_{R_1} + \mu_{R_2} + \cdots + \mu_{R_m} = \frac{m(m + n + 1)}{2} \tag{15.4}$$

La varianza de R_i fácilmente se calcula también que es $(m + n + 1)(m + n - 1)/12$. No obstante, como las R_i no son variables independientes, $V(W) \neq mV(R_i)$. Usando el hecho de que, para cualesquier dos enteros distintos a y b entre 1 y $m + n$ inclusive, $P(R_i = a, R_j = b) = 1/[(m + n)(m + n - 1)]$ (dos enteros se muestrean sin restitución), $\text{Cov}(R_i, R_j) = -(m + n + 1)/12$, que da

$$\sigma_W^2 = \sum_{i=1}^{m} V(R_i) + \sum\sum_{i \neq j} \text{Cov}(R_i, R_j) = \frac{mn(m + n + 1)}{12} \tag{15.5}$$

Es posible usar un teorema de límite central para concluir que cuando H_0 es verdadera, el *estadístico de prueba*

$$Z = \frac{W - m(m + n + 1)/2}{\sqrt{mn(m + n + 1)/12}}$$

tiene aproximadamente una distribución normal estándar. Este estadístico se usa en conjunción con los valores críticos z_α, $-z_\alpha$ y $\pm z_{\alpha/2}$ para pruebas de cola superior, cola inferior y dos colas, respectivamente.

Ejemplo 15.5 El artículo "Histamine Content in Sputum from Allergic and Non-Allergic Individuals", (*J. of Appl. Physiology*, 1969: 535–539) publica la siguiente información sobre el nivel de histamina en el esputo (μg/g de peso seco de esputo), para una muestra de 9 individuos clasificados como alérgicos y otra muestra de 13 individuos clasificados como no alérgicos:

Alérgicos	67.6	39.6	1651.0	100.0	65.9	1112.0	31.0	102.4	64.7				
No alérgicos	34.3	27.3	35.4	48.1	5.2	29.1	4.7	41.7	48.0	6.6	18.9	32.4	45.5

¿Esta información indica que hay una diferencia en el promedio verdadero de nivel de histamina en esputo entre alérgicos y no alérgicos? Vamos a seguir el ejemplo de los autores del artículo y usar la prueba Wilcoxon.

Como ambos tamaños muestrales exceden de 8, se usa la aproximación normal. La hipótesis nula es H_0: $\mu_1 - \mu_2 = 0$, y los rangos observados de las x_i son $r_1 = 18$, $r_2 = 11$, $r_3 = 22$, $r_4 = 19$, $r_5 = 17$, $r_6 = 21$, $r_7 = 7$, $r_8 = 20$, y $r_9 = 16$, de modo que $w = \sum r_i = 151$. La media y varianza de W están dadas por $\mu_W = 9(23)/2 = 103.5$ y $\sigma_W^2 = 9(13)(23)/12 = 224.25$. Entonces

$$z = \frac{151 - 103.5}{\sqrt{224.25}} = 3.17$$

La hipótesis alternativa es H_a: $\mu_1 - \mu_2 \neq 0$, de modo que al nivel .01 la H_0 es rechazada si $z \geq 2.58$ o $z \leq -2.58$. Como $3.17 \geq 2.58$, H_0 es rechazada y se concluye que hay una diferencia en el promedio verdadero de niveles de histamina en esputo. ■

Si hay empates en los datos, el numerador de Z todavía es apropiado, pero el denominador debe ser sustituido por la raíz cuadrada de la varianza ajustada

$$\sigma_W^2 = \frac{mn(m + n + 1)}{12} - \frac{mn}{12(m + n)(m + n - 1)} \sum(\tau_i - 1)(\tau_i)(\tau_i + 1) \qquad (15.6)$$

donde τ_i es el número de observaciones empatadas del i-ésimo conjunto de empates y la suma es sobre todos los conjuntos de enlaces. A menos que haya un gran número de empates, hay poca diferencia entre las ecuaciones (15.6) y (15.5).

Eficiencia de la prueba Wilcoxon de suma de rangos

Cuando las distribuciones que se muestrean son normales con $\sigma_1 = \sigma_2$ y por tanto tienen las mismas formas y dispersiones, se pueden usar la prueba t agrupada o la prueba Wilcoxon (la prueba t de dos muestras supone normalidad pero no varianzas iguales, de modo que las suposiciones que sirven de base a su uso son más restrictivas en un sentido y menos en otro que las suposiciones para la prueba Wilcoxon). En esta situación, la prueba t agrupada es mejor entre todas las pruebas posibles en el sentido de minimizar β para cualquier α fija. Sin embargo, un investigador nunca puede estar absolutamente seguro de que se satisfacen las suposiciones fundamentales. Por tanto, es importante preguntar (1) cuánto se pierde al usar la prueba de Wilcoxon, en lugar de la prueba t agrupada, cuando las distribuciones son normales con iguales varianzas y (2) cómo se compara W con T en situaciones que no sean normales.

El concepto de eficiencia de prueba se discutió en la sección previa en relación con una prueba t de una muestra y la prueba Wilcoxon de rango con signo. Los resultados para las pruebas de dos muestras son iguales que los de las pruebas de una muestra. Cuando se cumplen tanto normalidad como varianzas iguales, la prueba de la suma de rangos es

aproximadamente 95% de eficiente que la prueba *t* agrupada en muestras grandes. Es decir, la prueba *t* dará las mismas probabilidades de error que la prueba Wilcoxon que usa tamaños muestrales ligeramente más pequeños. Por otra parte, la prueba Wilcoxon siempre será al menos 86% de eficiente que la prueba *t* agrupada y puede ser más eficiente si las distribuciones fundamentales son, en gran medida, no normales, en especial con colas gruesas. La comparación de la prueba Wilcoxon con la prueba *t* de dos muestras (no agrupada) es menos clara. No se sabe que la prueba *t* sea la mejor prueba en ningún sentido, de modo que parece seguro concluir que mientras las distribuciones de población tengan formas y dispersiones semejantes, el comportamiento de la prueba Wilcoxon debe compararse en forma bastante favorable con la prueba *t* de dos muestras.

Por último, se observa que los cálculos β para la prueba Wilcoxon son muy difíciles. Esto es porque la distribución de W cuando H_0 es falsa depende no sólo de $\mu_1 - \mu_2$ sino también de las formas de las dos distribuciones. Para casi todas las distribuciones fundamentales, la distribución no nula de W es prácticamente insoluble. Ésta es la razón por la cual los expertos en estadística han creado una eficiencia de muestra grande (asintótica relativa) como medio de comparar pruebas. Con la capacidad de los modernos paquetes de software, otro método para el cálculo de β es realizar un experimento de simulación.

EJERCICIOS Sección 15.2 (10–16)

10. En un experimento para comparar la resistencia de pegamento de dos adhesivos diferentes, se utilizó cada adhesivo en cinco uniones de dos superficies y para cada unión se determinó la fuerza necesaria para separar las superficies. Para el adhesivo 1, los valores resultantes fueron 229, 286, 245, 299 y 250, mientras que las observaciones para el adhesivo 2 fueron 213, 179, 163, 247 y 225. Denótese con μ_i el promedio verdadero de resistencia de pegamento del adhesivo tipo *i*. Utilice la prueba Wilcoxon de suma de rangos al nivel .05 para probar H_0: $\mu_1 = \mu_2$ en función de H_a: $\mu_1 > \mu_2$.

11. El artículo "A Study of Wood Stove Particulate Emissions" (*J. of the Air Pollution Control Assoc.*, 1979: 724–728) publica la información siguiente sobre el tiempo de combustión (en horas) para muestras de roble y pino. Pruebe al nivel .05 para ver si hay alguna diferencia en el promedio verdadero de tiempo de combustión para los dos tipos de madera.

Roble	1.72	.67	1.55	1.56	1.42	1.23	1.77	.48
Pino	.98	1.40	1.33	1.52	.73	1.20		

12. Se ha hecho una modificación al proceso para producir cierto tipo de película de "tiempo cero" (película que empieza a revelarse tan pronto como se toma una foto). Debido a que la modificación implica un costo extra, será incorporada sólo si los datos muestrales indican firmemente que la modificación ha reducido el promedio verdadero de tiempo de revelado en más de 1 segundo. Si se supone que las distribuciones de tiempo de revelado difieren sólo con respecto al lugar, si es que difieren, utilice la prueba Wilcoxon de suma de rangos al nivel .05 con los datos del ejercicio 13 para probar las hipótesis apropiadas.

Proceso original	8.6	5.1	4.5	5.4	6.3	6.6	5.7	8.5
Proceso modificado	5.5	4.0	3.8	6.0	5.8	4.9	7.0	5.7

13. La información que sigue resultó de un experimento para comparar los efectos de la vitamina C, de jugo de naranja y de ácido ascórbico sintético, sobre la duración de odontoblastos en conejillos de Indias en un periodo de seis semanas ("The Growth of the Odontoblasts of the Incisor Tooth as a Criterion of the Vitamin C Intake of the Guinea Pig", *J. of Nutrition*, 1947: 491-504). Utilice la prueba Wilcoxon de suma de rangos al nivel .01 para determinar si el promedio verdadero de duración difiere para los dos tipos de ingesta de vitamina C. Calcule también un valor *P* apropiado.

Jugo de naranja	8.2	9.4	9.6	9.7	10.0	14.5
	15.2	16.1	17.6	21.5		
Ácido ascórbico	4.2	5.2	5.8	6.4	7.0	7.3
	10.1	11.2	11.3	11.5		

14. El artículo "Multimodal Versus Unimodal Instruction in a Complex Learning Environment" (*J. of Experimental Educ.*, 2002: 215–239) describe un experimento llevado a cabo para comparar el dominio de los estudiantes de un determinado software aprendido en dos formas diferentes. El primer método de aprendizaje (instrucción multimodal) implicó el uso de un manual visual. La segunda técnica (instrucción unimodal) empleó un manual de texto. Los siguientes son los resultados del aprendizaje para los dos grupos al final del experimento (la asignación a los grupos fue al azar):

Método 1:	44.85	46.59	47.60	51.08	52.20	56.87	57.03	57.07	60.35
	60.82	67.30	70.15	70.77	75.21	75.28	76.60	80.30	81.23
Método 2:	39.91	42.01	43.58	48.83	49.07	49.48	49.57	49.63	50.75
	51.95	56.54	57.40	57.60	61.16	64.55	65.31	68.59	72.40

¿Los datos sugieren que la puntuación promedio real depende de qué método de aprendizaje se utiliza?

15. El artículo "Measuring the Exposure of Infants to Tobacco Smoke" (*New Engl. J. of Med.*, 1984: 1075–1078) informa sobre un estudio en el que se tomaron varias medidas, tanto de una muestra aleatoria de infantes que habían estado expuestos al humo en una casa como de una muestra de infantes no expuestos. La información siguiente consta de observaciones en concentración de cotanina urinaria, metabolito importante de la nicotina (los valores son un subconjunto de la información original y se leyeron de una gráfica que apareció en el artículo). ¿La información sugiere que el promedio verdadero de nivel de cotanina es más alto, en más de 25, en infantes expuestos que en los no expuestos? Realice una prueba al nivel de significancia de .05.

No expuestos	8	11	12	14	20	43	111	
Expuestos	35	56	83	92	128	150	176	208

16. Reconsidere la situación descrita en el ejercicio 81 del capítulo 9 y la siguiente salida Minitab (la letra griega eta se emplea para denotar una mediana).

```
Intervalo de confianza y prueba de Mann-Whitney
buena              N = 8              Mediana = 0.540
pobre              N = 8              Mediana = 2.400
El punto estimado para ETA1-ETA2 es      −1.155
IC del 95.9% para ETA1-ETA2 es (−3.160, −0.409)
W = 41.0
Prueba de ETA1 = ETA2 contra ETA1 < ETA2 con signi-
ficancia de 0.0027
```

a. Verifique que el valor del estadístico de prueba de Minitab es correcto.

b. Realice una prueba apropiada de las hipótesis usando un nivel de significancia de .01.

15.3 Intervalos de confianza de distribución libre

El método empleado hasta aquí para construir un intervalo de confianza (IC) se puede describir como sigue: inicie con una variable aleatoria (Z, T, χ^2, F, u otras semejantes) que dependa del parámetro de interés y un enunciado de probabilidad que comprenda la variable, manipule las desigualdades del enunciado para despejar el parámetro entre puntos extremos aleatorios, y finalmente sustituya los valores calculados para variables aleatorias. Otro método general para obtener intervalos de confianza se aprovecha de una relación entre procedimientos de prueba e intervalos de confianza. Un intervalo de confianza de $100(1-\alpha)\%$ para un parámetro θ se puede obtener de una prueba de nivel α para H_0: $\theta = \theta_0$ en función de H_a: $\theta \neq \theta_0$. Este método se utilizará para deducir intervalos asociados con la prueba Wilcoxon de rango con signo y la prueba Wilcoxon de suma de rangos.

Para apreciar cómo se deducen los nuevos intervalos, reconsidere la prueba t para una muestra y el intervalo t. Suponga que una muestra aleatoria de $n = 25$ observaciones de una población normal da $\bar{x} = 100$, $s = 20$. Entonces un intervalo de confianza de 90% para μ es

$$\left(\bar{x} - t_{.05,24} \cdot \frac{s}{\sqrt{25}},\ \bar{x} + t_{.05,24} \cdot \frac{s}{\sqrt{25}}\right) = (93.16,\ 106.84) \qquad (15.7)$$

Ahora vamos a cambiar de marcha y probar hipótesis. Para H_0: $\mu = \mu_0$ contra H_a: $\mu \neq \mu_0$, la prueba t al nivel .10 especifica que H_0 debe ser rechazada si t es ≥ 1.711 o ≤ -1.711, donde

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{25}} = \frac{100 - \mu_0}{20/\sqrt{25}} = \frac{100 - \mu_0}{4} \qquad (15.8)$$

Considere ahora el valor nulo $\mu_0 = 95$. Entonces $t = 1.25$ de modo que H_0 no es rechazada. Del mismo modo, si $\mu_0 = 104$, entonces $t = -1$ por lo que de nuevo H_0 no es rechazada. No obstante, si $\mu_0 = 90$, entonces $t = 2.5$ y H_0 es rechazada y si $\mu_0 = 108$, entonces $t = -2$ y H_0 es de nuevo rechazada. Al considerar otros valores de μ_0 y la decisión resultante de cada uno, emerge el siguiente dato general: *todo número dentro del intervalo* (15.7) *especifica un valor para* μ_0 *para el que t de* (15.8) *lleva al no rechazo de* H_0, *mientras que todo número fuera del intervalo* (15.7) *corresponde a una t para la que* H_0 *es rechazada.* Es decir, para los valores fijos de n, $\bar{x}$, y s, el conjunto de todos los valores μ_0 para los cuales probar H_0: $\mu = \mu_0$ en función de H_a: $\mu \neq \mu_0$ resulta en no rechazo de H_0 es precisamente el intervalo (15.7).

PROPOSICIÓN

Suponga que se tiene un procedimiento de prueba de nivel α para probar H_0: $\theta = \theta_0$ en función de H_a: $\theta \neq \theta_0$. Para valores muestrales fijos, denótese con A el conjunto de todos los valores θ_0 para los cuales H_0 no es rechazada. Entonces A es un intervalo de confianza de $100(1 - \alpha)\%$ para θ.

En realidad hay ejemplos patológicos en los que el conjunto A definido en la proposición no es un intervalo de valores θ, sino que es el complemento de un intervalo o algo todavía más extraño. Para ser más precisos, se debería sustituir realmente la idea de un intervalo de confianza con la de un conjunto de confianza. En los casos de interés aquí, el conjunto A resulta ser un intervalo.

El intervalo Wilcoxon de rango con signo

Para probar H_0: $\mu = \mu_0$ contra H_a: $\mu \neq \mu_0$ usando la prueba Wilcoxon de rango con signo, donde μ es la media de una distribución simétrica continua, los valores absolutos $|x_1 - \mu_0|, \ldots, |x_n - \mu_0|$ se ordenan de menor a mayor, con el más pequeño recibiendo rango 1 y, el mayor, rango n. A cada rango se da entonces el signo de su $x_i - \mu_0$ asociada, y el estadístico de prueba es la suma de los rangos con signo positivo. La prueba de dos colas rechaza H_0 si s_+ es $\geq c$ o $\leq n(n + 1)/2 - c$, donde c se obtiene de la tabla A.13 del apéndice una vez especificado el nivel deseado de significancia α. Para $x_1, \ldots, x_n$ fijas, el intervalo de rango con signo $100(1 - \alpha)\%$ estará formado por todas las μ_0 para las que H_0: $\mu = \mu_0$ no es rechazada al nivel α. Para identificar este intervalo, es conveniente expresar el estadístico de prueba S_+ en otra forma.

$$S_+ = \text{el número de promedios por pares } (X_i + X_j)/2 \text{ con } i \leq j \text{ que son } \geq \mu_0 \quad (15.9)$$

Esto es, si se promedia cada una de las x_j de la lista con cada una de las x_i a su izquierda, incluyendo $(x_j + x_j)/2$ (que es sólo x_j), y se cuenta el número de estos promedios que sean $\geq \mu_0$, resulta s_+. Al moverse de izquierda a derecha en la lista de valores muestrales, simplemente se está promediando cada par de observaciones de la muestra [otra vez incluyendo $(x_j + x_j)/2$] exactamente una vez, de modo que el orden en que se ponen en lista las observaciones antes de promediarse no es importante. La equivalencia de los dos métodos para calcular s_+ no es difícil de verificar. El número de promedios por pares es $\binom{n}{2} + n$ (el primer término se debe a promediar diferentes observaciones y el segundo a promediar cada x_i consigo misma), que es igual a $n(n + 1)/2$. Si demasiados, o demasiado pocos, de estos promedios por pares son $\geq \mu_0$, H_0 es rechazada.

Ejemplo 15.6

Las siguientes observaciones son valores de ritmo metabólico cerebral para monos rhesus: $x_1 = 4.51$, $x_2 = 4.59$, $x_3 = 4.90$, $x_4 = 4.93$, $x_5 = 6.80$, $x_6 = 5.08$, $x_7 = 5.67$. Los 28 promedios por pares son, en orden creciente,

4.51	4.55	4.59	4.705	4.72	4.745	4.76	4.795	4.835	4.90
4.915	4.93	4.99	5.005	5.08	5.09	5.13	5.285	5.30	5.375
5.655	5.67	5.695	5.85	5.865	5.94	6.235	6.80		

Los primeros pocos y los pocos últimos de éstos se presentan en un eje de mediciones en la figura 15.2.

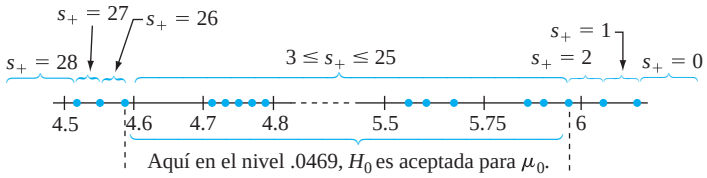


Figura 15.2 Gráfica de los datos para el ejemplo 15.6

Como S_+ es una variable aleatoria discreta, $\alpha = .05$ no se puede obtener de manera exacta. La región de rechazo $\{0, 1, 2, 26, 27, 28\}$ tiene $\alpha = .046$, que es tan cercano como es posible a .05, de modo que el nivel es aproximadamente .05. De este modo, si el número de promedios por pares $\geq \mu_0$ está entre 3 y 25, inclusive, H_0 no es rechazada. De la figura 15.2 el intervalo de confianza (aproximado) de 95% para μ es (4.59, 5.94). ■

En general, una vez que los promedios por pares sean ordenados de menor a mayor, los puntos extremos del intervalo Wilcoxon son dos de los promedios "extremos". Para expresar esto con precisión, denótese por $\bar{x}_{(1)}$ el promedio por pares más pequeño, el siguiente más pequeño por $\bar{x}_{(2)}$, . . . , y el más grande por $\bar{x}_{(n(n+1)/2)}$.

PROPOSICIÓN

Si la prueba Wilcoxon de rango con signo de nivel α para H_0: $\mu = \mu_0$ en función de H_a: $\mu \neq \mu_0$ es para rechazar H_0 si $s_+ \geq c$ o $s_+ \leq n(n+1)/2 - c$, entonces un intervalo de confianza $100(1 - \alpha)\%$ para μ es

$$(\bar{x}_{(n(n+1)/2-c+1)}, \bar{x}_{(c)}) \qquad (15.10)$$

En otras palabras, el intervalo se extiende desde el d-ésimo más pequeño promedio por pares al d-ésimo promedio más grande, donde $d = n(n+1)/2 - c + 1$. La tabla A.15 del apéndice da los valores de c que corresponden a los niveles de confianza acostumbrados para $n = 5, 6, \ldots, 25$.

Ejemplo 15.7 (Continuación del ejemplo 15.6)

Para $n = 7$, se obtiene un intervalo de 89.1% (aproximadamente 90%) con el uso de $c = 24$ (porque la región de rechazo (0, 1, 2, 3, 4, 24, 25, 26, 27, 28) tiene $\alpha = .109$). El intervalo es $(\bar{x}_{(28-24+1)}, \bar{x}_{(24)}) = (\bar{x}_{(5)}, \bar{x}_{(24)}) = (4.72, 5.85)$, que se extiende desde el quinto promedio más pequeño al quinto promedio por pares más grande. ■

La deducción del intervalo dependía de tener una sola muestra de una distribución simétrica continua con media (mediana) μ. Cuando la información es pareada, el intervalo construido desde las diferencias $d_1, d_2, \ldots, d_n$ es un intervalo de confianza para la diferencia media (mediana) μ_D. En este caso, no es necesario suponer la simetría de las distribuciones X y Y; mientras las distribuciones X y Y tengan la misma forma, la distribución $X - Y$ será simétrica, por lo que sólo se requiere continuidad.

Para $n > 20$, la aproximación de muestra grande a la prueba Wilcoxon basada en estandarizar S_+ da una aproximación a c en (15.10). El resultado [para un intervalo $100(1 - \alpha)\%$] es

$$c \approx \frac{n(n+1)}{4} + z_{\alpha/2}\sqrt{\frac{n(n+1)(2n+1)}{24}}$$

La eficiencia del relativo intervalo Wilcoxon con respecto al intervalo t es aproximadamente igual que el relativo para la prueba Wilcoxon con respecto a la prueba t. En particular, para muestras grandes cuando la población fundamental es normal, el intervalo Wilcoxon tenderá a ser un poco más largo que el intervalo t, pero si la población es no nor-

mal en suficiente grado (simétrica pero con colas gruesas), entonces el intervalo Wilcoxon tenderá a ser mucho más corto que el intervalo *t*.

El intervalo Wilcoxon de suma de rangos

La prueba Wilcoxon de suma de rangos para probar H_0: $\mu_1 - \mu_2 = \Delta_0$ se realiza al combinar primeramente $(X_i - \Delta_0)$ y las Y_j en una muestra de tamaño $m + n$, y ordenarlas de menor (rango 1) a mayor (rango $m + n$). El estadístico de prueba W es entonces la suma de los rangos de las $(X_i - \Delta_0)$. Para la alternativa de dos lados, H_0 es rechazada si w es demasiado pequeña o demasiado grande.

Para obtener el intervalo de confianza asociado para las x_i y las y_j fijas, se debe determinar el conjunto de todos los valores Δ_0 para los que H_0 no es rechazada. Esto es más fácil de hacer si primero se expresa el estadístico de prueba en una forma ligeramente diferente. El valor más pequeño posible de W es $m(m + 1)/2$, correspondiente a cada $(X_i - \Delta_0)$ menor que cada Y_j, y hay mn diferencias de la forma $(X_i - \Delta_0) - Y_j$. Un poco de manipulación da

$$\begin{aligned} W &= [\text{número de}(X_i - Y_j - \Delta_0) \geq 0] + \frac{m(m+1)}{2} \\ &= [\text{número de } (X_i - Y_j) \geq \Delta_0] + \frac{m(m+1)}{2} \end{aligned} \tag{15.11}$$

Entonces, el rechazo de H_0 si el número de las $(x_i - y_j) \geq \Delta_0$ es demasiado pequeño o demasiado grande equivale a rechazar H_0 para w pequeña o grande.

La expresión (15.11) sugiere que se calcule $x_i - y_j$ para cada i y j y se ordenen estas mn diferencias de menor a mayor. Entonces, si el valor nulo Δ_0 no es más pequeño que la mayor parte de las diferencias ni mayor que casi todas, H_0: $\mu_1 - \mu_2 = \Delta_0$ no es rechazada. La variación de Δ_0 muestra ahora que un intervalo de confianza para $\mu_1 - \mu_2$ tendrá como su punto extremo inferior una de las $(x_i - y_j)$, y análogamente para el punto extremo superior.

PROPOSICIÓN

Sean $x_1, \ldots, x_m$ y $y_1, \ldots, y_n$ los valores observados en dos muestras independientes de distribuciones continuas que difieren sólo en ubicación (y no en forma). Con $d_{ij} = x_i - y_j$ y las diferencias ordenadas denotadas por $d_{ij(1)}, d_{ij(2)}, \ldots, d_{ij(mn)}$, la forma general de un intervalo de confianza $100(1 - \alpha)\%$ para $\mu_1 - \mu_2$ es

$$(d_{ij(mn-c+1)}, d_{ij(c)}) \tag{15.12}$$

donde c es la constante crítica para la prueba Wilcoxon de suma de rangos de nivel α de dos colas.

Nótese que la forma del intervalo Wilcoxon de suma de rangos (15.12) es muy semejante al intervalo Wilcoxon de rango con signo (15.10); (15.10) emplea promedios por pares de una sola muestra, mientras que (15.12) usa diferencias por pares de dos muestras. La tabla A.16 del apéndice da valores de c para valores seleccionados de m y n.

Ejemplo 15.8 El artículo “Some Mechanical Properties of Impregnated Bark Board” (*Forest Products J.* 1977: 31–38) publica la siguiente información sobre la máxima resistencia a la compresión (por pulgada cuadrada), para una muestra de tabla de corteza de impregnación epóxica y para una muestra de tabla de corteza impregnada con otro polímero:

Epóxica (de *x*)	10,860	11,120	11,340	12,130	14,380	13,070
Otro (de *y*)	4590	4850	6510	5640	6390	

Obtenga un intervalo de confianza de 95% para el promedio verdadero de diferencia en resistencia a la compresión entre la tabla de impregnación epóxica y la tabla con el otro tipo.

De la tabla A.16 del apéndice, puesto que el tamaño muestral más pequeño es 5 y el tamaño muestral más grande es 6, $c = 26$ para un nivel de confianza de aproximadamente 95%. Las d_{ij} aparecen en la tabla 15.5. Las cinco d_{ij} más pequeñas $[d_{ij(1)}, \ldots, d_{ij(5)}]$ son 4350, 4470, 4610, 4730 y 4830 y las cinco más grandes son (en orden descendente 9790, 9530, 8740, 8480 y 8220. Entonces el intervalo de confianza es $(d_{ij(5)}, d_{ij(26)}) = \ = (4830, 8220)$.

Tabla 15.5 Diferencias para el intervalo de suma de rangos del ejemplo 15.8

	d_{ij}	y_j 4590	4850	5640	6390	6510
	10,860	6270	6010	5220	4470	4350
	11,120	6530	6270	5480	4730	4610
x_i	**11,340**	6750	6490	5700	4950	4830
	12,130	7540	7280	6490	5740	5620
	13,070	8480	8220	7430	6680	6560
	14,380	9790	9530	8740	7990	7870

Cuando m y n son grandes, el estadístico de prueba Wilcoxon tiene aproximadamente una distribución normal. Ésta se puede usar para deducir una aproximación de muestra grande para el valor c en el intervalo (15.12). El resultado es

$$c \approx \frac{mn}{2} + z_{\alpha/2}\sqrt{\frac{mn(m+n+1)}{12}} \qquad (15.13)$$

Al igual que con el intervalo de rango con signo, el intervalo de prueba de rangos (15.12) es muy eficiente con respecto al intervalo t; en muestras grandes, (15.12) tenderá a ser sólo un poco más larga que el intervalo t, cuando las poblaciones fundamentales son normales, y puede ser considerablemente más corta que el intervalo t si las poblaciones fundamentales tienen colas más gruesas que las poblaciones normales.

EJERCICIOS Sección 15.3 (17–22)

17. El artículo "The Lead Content and Acidity of Christchurch Precipitation" (*New Zealand J. of Science*, 1980: 311–312) publica los datos siguientes sobre la concentración de plomo (μg/L) en muestras tomadas durante lluvias diferentes en verano: 17.0, 21.4, 30.6, 5.0, 12.2, 11.8, 17.3 y 18.8. Suponiendo que la distribución del contenido de plomo sea simétrica, utilice el intervalo Wilcoxon de rango con signo para obtener un intervalo de confianza de 95% para μ.

18. Calcule un intervalo de rango con signo de 99% para el promedio verdadero de pH, μ (suponiendo simetría) usando los datos del ejercicio 15.3. [*Sugerencia:* trate de calcular sólo los promedios apareados que tengan valores relativamente pequeños o grandes (más que todos los 105 promedios).]

19. Se realizó un experimento para comparar las capacidades de dos diferentes disolventes para extraer la creosota impregnada en la prueba de los registros. Cada uno de los ocho registros se dividió en dos segmentos, a continuación, un segmento fue seleccionado al azar para la aplicación del primer solvente, con el otro segmento receptor del segundo solvente.

Log	1	2	3	4	5	6	7	8
Solvente 1	3.92	3.79	3.70	4.08	3.87	3.95	3.55	3.76
Solvente 2	4.25	4.20	4.41	3.89	4.39	3.75	4.20	3.90

Calcular un intervalo de confianza con un nivel de confianza de aproximadamente el 95% de la diferencia entre la cantidad promedio real extraída por medio del primer solvente y la cantidad promedio real extraída por medio del segundo.

20. Las siguientes observaciones son cantidades de emisiones de hidrocarburos, que resultan del desgaste en pavimento de neumáticos de banda en diagonal bajo una carga de 522 kg infla-

das a 228 kPa y corriendo a 64 km/h durante 6 horas ("Characterization of Tire Emissions Using an Indoor Test Facility", *Rubber Chemistry and Technology*, 1978: 7–25): .045, .117, .062, y .072. ¿Qué niveles de confianza se pueden alcanzar para este tamaño muestral si se usa el intervalo de rango con signo? Seleccione un nivel de confianza apropiado y calcule el intervalo.

21. Calcule el intervalo de confianza de suma de rangos de 90% para $\mu_1 - \mu_2$ usando los datos del ejercicio 10.

22. Calcule un intervalo de confianza de 99% para $\mu_1 - \mu_2$ usando los datos del ejercicio 11.

15.4 ANOVA de distribución libre

El modelo ANOVA de un solo factor del capítulo 10, para comparar I medias de una población o tratamiento, significa suponer que para $i = 1, 2, \ldots, I$, una muestra aleatoria de tamaño J_i se extrajo de una población normal con media μ_i y varianza σ^2. Esto se puede escribir como

$$X_{ij} = \mu_i + \epsilon_{ij} \quad j = 1, \ldots, J_i;\ i = 1, \ldots, I \tag{15.14}$$

donde las ϵ_{ij} están independiente y normalmente distribuidas con media cero y varianza σ^2. Aun cuando la suposición de normalidad se requirió para la validez de la prueba F descrita en el capítulo 10, el siguiente procedimiento para probar la igualdad de las μ_i requiere sólo que las ϵ_{ij} que tengan la misma distribución continua.

La prueba Kruskal-Wallis

Sea $N = \sum J_i$, el número total de observaciones del conjunto de datos, y suponga que se ordenan todas las N observaciones de 1 (la X_{ij} más pequeña) a N (la X_{ij} más grande). Cuando H_0: $\mu_1 = \mu_2 = \cdots = \mu_I$ es verdadera, las N observaciones provienen todas de la misma distribución, en cuyo caso todas las posibles asignaciones de los rangos 1, 2, . . . , N a las I muestras son igualmente probables y se espera que los rangos se mezclen en estas muestras. Pero, si H_0 es falsa, entonces algunas muestras estarán formadas sobre todo de observaciones que tengan rangos pequeños en la muestra combinada, mientras que otras estarán formadas principalmente de observaciones que tengan rangos grandes. En forma más específica, si R_{ij} denota el rango de X_{ij} entre las N observaciones y $R_{i\cdot}$ y $\overline{R}_{i\cdot}$ denotan, respectivamente, el total y promedio de los rangos de la i-ésima muestra, entonces cuando H_0 es verdadera

$$E(R_{ij}) = \frac{N+1}{2} \qquad E(\overline{R}_{i\cdot}) = \frac{1}{J_i}\sum_j E(R_{ij}) = \frac{N+1}{2}$$

El estadístico de prueba Kruskal-Wallis es una medida de la magnitud a la que las $\overline{R}_{i\cdot}$ se desvían de su valor esperado común $(N+1)/2$, y H_0 es rechazada si el valor calculado del estadístico indica una discrepancia demasiado grande entre promedios de rango observados y esperados

ESTADÍSTICO DE PRUEBA

$$K = \frac{12}{N(N+1)}\sum_{j=1}^{I} J_i\left(\overline{R}_{i\cdot} - \frac{N+1}{2}\right)^2$$
$$= \frac{12}{N(N+1)}\sum_{i=1}^{I}\frac{R_{i\cdot}^2}{J_i} - 3(N+1) \tag{15.15}$$

La segunda expresión para K es la fórmula computacional; comprende los totales de rango ($R_{i\cdot}$) más que los promedios y requiere sólo una resta.

Si H_0 es rechazada cuando $k \geq c$, entonces c debe seleccionarse de modo que la prueba tenga nivel α. Esto es, c debe ser el valor crítico de cola superior de la distribución de K cuando H_0 sea verdadera. Bajo H_0, cada asignación posible de los rangos de las I muestras es igualmente probable, de modo que en teoría todas las asignaciones se pueden enumerar, el valor de K determinarse para cada una, y obtenerse la distribución nula al contar el número de veces que se presente cada uno de los valores de K. Es evidente que este cálculo es tedioso, de manera que aun cuando hay tablas de la distribución nula exacta y valores críticos para valores pequeños de las J_i, se usará la siguiente aproximación de "muestra grande".

PROPOSICIÓN

Cuando H_0 es verdadera y

$$I = 3 \quad J_i \geq 6 \quad (i = 1, 2, 3)$$

o

$$I > 3 \quad J_i \geq 5 \quad (i = 1, \ldots, I)$$

entonces K tiene aproximadamente una distribución chi cuadrada con grado de libertad $I - 1$. Esto implica que una prueba con nivel de significancia aproximado α rechaza H_0 si $k \geq \chi^2_{\alpha, I-1}$.

Ejemplo 15.9 Las observaciones siguientes (tabla 15.6), sobre el índice de rigidez axial, resultaron de un estudio de armazones conectadas con placas metálicas en las que se usaron cinco diferentes longitudes de placa: 4, 6, 8, 10 y 12 pulgadas ("Modeling Joints Made with Light-Gauge Metal Connector Plates", *Forest Products J.*, 1979: 39–44).

Tabla 15.6 Datos y rangos para el ejemplo 15.9

	$i = 1$ (4″):	309.2	309.7	311.0	316.8	326.5	349.8	409.5		
	$i = 2$ (6″):	331.0	347.2	348.9	361.0	381.7	402.1	404.5		
	$i = 3$ (8″):	351.0	357.1	366.2	367.3	382.0	392.4	409.9		
	$i = 4$ (10″):	346.7	362.6	384.2	410.6	433.1	452.9	461.4		
	$i = 5$ (12″):	407.4	410.7	419.9	441.2	441.8	465.8	473.4		
									$r_{i\cdot}$	$\bar{r}_{i\cdot}$
Rangos	$i = 1$:	1	2	3	4	5	10	24	49	7.00
	$i = 2$:	6	8	9	13	17	21	22	96	13.71
	$i = 3$:	11	12	15	16	18	20	25	117	16.71
	$i = 4$:	7	14	19	26	29	32	33	160	22.86
	$i = 5$:	23	27	28	30	31	34	35	208	29.71

El valor calculado de K es

$$k = \frac{12}{35(36)}\left[\frac{(49)^2}{7} + \frac{(96)^2}{7} + \frac{(117)^2}{7} + \frac{(160)^2}{7} + \frac{(208)^2}{7}\right] - 3(36)$$

$$= 20.21$$

Al nivel .01, $\chi^2_{.01,4} = 13.277$, y como $20.12 \geq 13.277$, H_0 es rechazada y se concluye que la rigidez axial esperada depende de la longitud de las placas. ■

Prueba de Friedman para un experimento de bloque aleatorizado

Suponga que $X_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$, donde α_i es el i-ésimo efecto de tratamiento, β_j es el j-ésimo efecto de bloque, y las ϵ_{ij} se extraen de manera independiente de la misma distribución continua (pero no necesariamente normal). Entonces para probar H_0: $\alpha_1 = \alpha_2 = \cdots = \alpha_I = 0$, la hipótesis nula de efectos sin tratamiento, primero se ordenan las observaciones en forma separada de 1 a I dentro de cada bloque, y luego se calcula el promedio de rango $\bar{r}_{i\cdot}$ para cada uno de los I tratamientos. Cuando H_0 es verdadera, las $\bar{r}_{i\cdot}$ deben ser cercanas entre sí, puesto que dentro de cada bloque todas las $I!$ asignaciones de rangos a tratamientos son igualmente probables. El estadístico de prueba de Friedman mide la discrepancia entre el valor esperado $(I + 1)/2$ de cada promedio de rango y las $\bar{r}_{i\cdot}$.

ESTADÍSTICO DE PRUEBA

$$F_r = \frac{12J}{I(I+1)} \sum_{i=1}^{I} \left(\bar{R}_{i\cdot} - \frac{I+1}{2} \right)^2 = \frac{12}{IJ(I+1)} \sum R_{i\cdot}^2 - 3J(I+1)$$

Al igual que con la prueba Kruskal-Wallis, la prueba de Friedman rechaza H_0 cuando el valor calculado del estadístico de prueba es demasiado grande. Para los casos $I = 3$, $J = 2, \ldots, 15$ e $I = 4$, $J = 2, \ldots, 8$, el libro de Lehmann (vea bibliografía del capítulo) da los valores críticos de cola superior para la prueba. De modo alternativo, para valores pares moderados de J, el estadístico de prueba F_r tiene aproximadamente una distribución chi cuadrada con grado de libertad $I - 1$ cuando H_0 es verdadera, de modo que H_0 puede ser rechazada si $f_r \geq \chi^2_{\alpha, I-1}$.

Ejemplo 15.10 El artículo "Physiological Effects During Hypnotically Requested Emotions" (*Psychosomatic Med.*, 1963: 334–343) publica la siguiente información (tabla 15.7) sobre el potencial en la piel (mV) cuando las emociones de temor, alegría, depresión, y calma se solicitaron de cada uno de ocho sujetos.

Tabla 15.7 Datos y rangos para el ejemplo 15.10

	Bloques (sujetos)									
x_{ij}	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**		
Temor	23.1	57.6	10.5	23.6	11.9	54.6	21.0	20.3		
Alegría	22.7	53.2	9.7	19.6	13.8	47.1	13.6	23.6		
Depresión	22.5	53.7	10.8	21.1	13.7	39.2	13.7	16.3		
Calma	22.6	53.1	8.3	21.6	13.3	37.0	14.8	14.8		
Rangos	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	$r_{i\cdot}$	$r_{i\cdot}^2$
Temor	4	4	3	4	1	4	4	3	27	729
Alegría	3	2	2	1	4	3	1	4	20	400
Depresión	1	3	4	2	3	2	2	2	19	361
Calma	2	1	1	3	2	1	3	1	14	196
										1686

Así,

$$f_r = \frac{12}{4(8)(5)}(1686) - 3(8)(5) = 6.45$$

Al nivel .05, $\chi^2_{.05,3} = 7.815$, y como $6.45 < 7.815$, H_0 no es rechazada. No hay evidencia de que el promedio de potencial en la piel dependa de qué emoción se pida. ■

El libro de Myles Hollander y Douglas Wolfe (vea bibliografía del capítulo) examina múltiples procedimientos de comparación asociados con las pruebas Kruskal-Wallis y Friedman, así como otros aspectos de ANOVA de distribución libre.

EJERCICIOS Sección 15.4 (23–27)

23. La información siguiente se refiere a la concentración del isótopo radiactivo estroncio-90, obtenida en muestras de leche de cinco lecherías seleccionadas al azar en cuatro regiones diferentes.

	1	6.4	5.8	6.5	7.7	6.1
Región	**2**	7.1	9.9	11.2	10.5	8.8
	3	5.7	5.9	8.2	6.6	5.1
	4	9.5	12.1	10.3	12.4	11.7

Pruebe al nivel .10 para ver si el promedio verdadero de la concentración de estroncio-90 difiere en al menos dos de las regiones.

24. El artículo "Production of Gaseous Nitrogen in Human Steady-State Conditions" (*J. of Applied Physiology*, 1972: 155–159) publica las siguientes observaciones sobre la cantidad de nitrógeno espirado (en litros) bajo cuatro regímenes de dieta: (1) ayuno, (2) 23% de proteínas, (3) 32% de proteínas, y (4) 67% de proteínas. Utilice la prueba Kruskal-Wallis al nivel .05 para probar la igualdad de las μ_i correspondientes.

1.	4.079	4.859	3.540	5.047	3.298
2.	4.368	5.668	3.752	5.848	3.802
3.	4.169	5.709	4.416	5.666	4.123
4.	4.928	5.608	4.940	5.291	4.674
1.	4.679	2.870	4.648	3.847	
2.	4.844	3.578	5.393	4.374	
3.	5.059	4.403	4.496	4.688	
4.	5.038	4.905	5.208	4.806	

25. Los datos siguientes sobre el nivel de hidrocortisona se publicaron en el artículo "Cortisol, Cortisone, and 11-Deoxycortisol Levels in Human Umbilical and Maternal Plasma in Relation to the Onset of Labor" (*J. of Obstetric Gynaecology of the British Commonwealth*, 1974: 737–745). Los sujetos experimentales fueron mujeres cuyos bebés nacieron con gestación de entre 38 y 42 semanas. Las mujeres del grupo 1 eligieron dar a luz por operación cesárea antes de principiar el trabajo de parto y las del grupo 2 dieron a luz por cesárea de emergencia durante un trabajo de parto inducido, y las del grupo 3 experimentaron un trabajo de parto espontáneo. Utilice la prueba Kruskal-Wallis al nivel .05 para probar la igualdad de las tres medias poblacionales.

Grupo 1	262	307	211	323	454	339
	304	154	287	356		
Grupo 2	465	501	455	355	468	362
Grupo 3	343	772	207	1048	838	687

26. En una prueba para determinar si un suelo tratado previamente con pequeñas cantidades de Basic-H hace que el suelo sea más permeable al agua, unas muestras de suelo se dividieron en bloques y cada uno de éstos recibió los cuatro tratamientos bajo estudio. Los tratamientos fueron (*A*) agua con .001% de Basic-H hasta cubrirla con tierra de control, (*B*) agua sin Basic-H en tierra de control, (*C*) agua con Basic-H cubierta con tierra tratada previamente con Basic-H y (*D*) agua sin Basic-H sobre tierra tratada previamente con Basic-H. Pruebe a un nivel .01 para ver si hay algún efecto debido a los diversos tratamientos.

	Bloques				
	1	**2**	**3**	**4**	**5**
A	37.1	31.8	28.0	25.9	25.5
B	33.2	25.3	20.2	20.3	18.3
C	58.9	54.2	49.2	47.9	38.2
D	56.7	49.6	46.4	40.9	39.4
	6	**7**	**8**	**9**	**10**
A	25.3	23.7	24.4	21.7	26.2
B	19.3	17.3	17.0	16.7	18.3
C	48.8	47.8	40.2	44.0	46.4
D	37.1	37.5	39.6	35.1	36.5

27. En un experimento para estudiar la forma en la que diferentes anestésicos afectan la concentración de epinefrina en plasma, se seleccionaron 10 perros y se les midió la concentración cuando estaban bajo el influjo de isoflurano, halotano y ciclopropano anestésicos ("Sympathoadrenal and Hemodynamic Effects of Isoflurane, Halothane, and Cyclopropane in Dogs", *Anesthesiology*, 1974: 465–470). Pruebe al nivel .05 para ver si hay un efecto anestésico en la concentración.

	Perro				
	1	**2**	**3**	**4**	**5**
Isoflurano	.28	.51	1.00	.39	.29
Halotano	.30	.39	.63	.38	.21
Ciclopropano	1.07	1.35	.69	.28	1.24
	6	**7**	**8**	**9**	**10**
Isoflurano	.36	.32	.69	.17	.33
Halotano	.88	.39	.51	.32	.42
Ciclopropano	1.53	.49	.56	1.02	.30

EJERCICIOS SUPLEMENTARIOS (28–36)

28. El artículo "Effects of a Rice-Rich Versus Potato-Rich Diet on Glucose, Lipoprotein, and Cholesterol Metabolism in Noninsulin-Dependent Diabetics" (*Amer. J. of Clinical Nutr.*, 1984: 598–606) da la información siguiente sobre la cantidad de síntesis de colesterol para ocho sujetos diabéticos. A éstos se les alimentó con una dieta estandarizada con papas o arroz como fuente principal de carbohidratos. Los participantes recibieron ambas dietas durante periodos especificados, con la cantidad de síntesis de colesterol (mmol/día) medida al final de cada periodo de dieta. El análisis presentado en este artículo utilizó una prueba de distribución libre. Use el lector esta prueba con nivel de significancia .05 para determinar si la media verdadera de la cantidad de síntesis de colesterol difiere de modo importante para las dos fuentes de carbohidratos.

Cantidad de síntesis de colesterol

Sujeto	1	2	3	4	5	6	7	8
Papas	1.88	2.60	1.38	4.41	1.87	2.89	3.96	2.31
Arroz	1.70	3.84	1.13	4.97	.86	1.93	3.36	2.15

29. Las tácticas de venta de alta presión o los vendedores de puerta en puerta pueden ser muy agresivos. Numerosas personas sucumben a esas tácticas, firman acuerdos de compra y posteriormente se lamentan de sus acciones. A mediados de la década de 1970, la Federal Trade Commission implantó reglamentos que aclaran y amplían los derechos de compradores para cancelar dichos acuerdos. La información siguiente es un subconjunto de la información dada en el artículo "Evaluating the FTC Cooling-Off Rule" (*J. of Consumer Affairs*, 1977: 101–106). Las observaciones individuales son porcentajes de cancelación para cada uno de nueve vendedores durante cada uno de 4 años. Utilice una prueba apropiada al nivel .05 para ver si el promedio verdadero de la cantidad de cancelaciones depende del año.

	Vendedor								
	1	2	3	4	5	6	7	8	9
1973	2.8	5.9	3.3	4.4	1.7	3.8	6.6	3.1	0.0
1974	3.6	1.7	5.1	2.2	2.1	4.1	4.7	2.7	1.3
1975	1.4	.9	1.1	3.2	.8	1.5	2.8	1.4	.5
1976	2.0	2.2	.9	1.1	.5	1.2	1.4	3.5	1.2

30. La información dada sobre la concentración de fósforo en tierra superficial, para cuatro tratamientos diferentes de suelos, apareció en el artículo "Fertilisers for Lotus and Clover Establishment on a Sequence of Acid Soils on the East Otago Uplands" (*N. Zeal. J. of Exptl. Ag.*, 1984: 119–129). Utilice un procedimiento de distribución libre para probar la hipótesis nula de no diferencia en la media verdadera de concentración de fósforo (mg/g) para los cuatro tratamientos de suelos.

Tratamiento						
	I	8.1	5.9	7.0	8.0	9.0
	II	11.5	10.9	12.1	10.3	11.9
	III	15.3	17.4	16.4	15.8	16.0
	IV	23.0	33.0	28.4	24.6	27.7

31. Consulte la información del ejercicio 30 y calcule un intervalo de confianza de 95% para la diferencia entre el promedio verdadero de concentraciones para los tratamientos II y III.

32. El estudio publicado en "Gait Patterns During Free Choice Ladder Ascents" (*Human Movement Sci.*, 1983: 187–195) fue motivado por publicidad relativa al mayor porcentaje de accidentes para individuos que suben por escaleras. De acuerdo con instrucciones especificadas se utilizaron varias formas diferentes de pasos por personas que trepaban por una escalera recta portátil. Se dan tiempos de ascenso para siete sujetos que usaron un paso lateral y seis sujetos que usaron un paso diagonal de cuatro tiempos.

Lateral	.86	1.31	1.64	1.51	1.53	1.39	1.09
Diagonal	1.27	1.82	1.66	.85	1.45	1.24	

a. Realice una prueba usando $\alpha = .05$ para ver si la información sugiere cualquier diferencia en el promedio verdadero de tiempos de ascenso para los dos pasos.

b. Calcule un intervalo de confianza de 95% para la diferencia entre el promedio verdadero de tiempos de paso.

33. La **prueba de signo** es un procedimiento muy sencillo para probar hipótesis acerca de una mediana poblacional que supone sólo que la distribución fundamental es continua. Para ilustrar, considere la siguiente muestra de 20 observaciones sobre la vida útil de un componente (h):

1.7	3.3	5.1	6.9	12.6	14.4	16.4
24.6	26.0	26.5	32.1	37.4	40.1	40.5
41.5	72.4	80.1	86.4	87.5	100.2	

Se desea probar H_0: $\tilde{\mu} = 25.0$ en función de H_a: $\tilde{\mu} > 25.0$. El estadístico de prueba es Y = el número de observaciones que exceden de 25.

a. Considere rechazar H_0 si $Y \geq 15$. ¿Cuál es el valor de α (la probabilidad de un error tipo I) para esta prueba? [*Sugerencia:* considere un "éxito" como una vida útil que exceda de 25.0. Entonces Y es el número de éxitos de la muestra.] ¿Qué clase de distribución tiene Y cuando $\tilde{\mu} = 25.0$?

b. ¿Qué región de rechazo de la forma $Y \geq c$ especifica una prueba con un nivel de significancia tan cercano a .05 como sea posible? Utilice esta región para efectuar la prueba para la información dada.

[*Nota*: el estadístico de prueba es el número de diferencias $X_i - 25$ que tiene signos positivos, por ello el nombre de *prueba de signo.*]

34. Consulte el ejercicio 33 y considere un intervalo de confianza asociado con la prueba de signo: el **intervalo de signo.** Las hipótesis relevantes son ahora H_0: $\tilde{\mu} = \tilde{\mu}_0$ en función de H_a: $\tilde{\mu} \neq \tilde{\mu}_0$. Úsese la siguiente región de rechazo: $Y \geq 15$ o $Y \leq 5$.

a. ¿Cuál es el nivel de significancia para esta prueba?

b. El intervalo de confianza constará de todos los valores $\tilde{\mu}_0$ para los cuales H_0 no es rechazada. Determine el intervalo de confianza para la información dada y exprese el nivel de confianza.

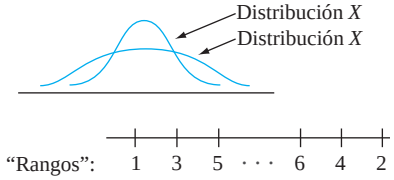
35. Suponga que desea probar

H_0: las distribuciones X y Y son idénticas

contra

H_a: la distribución X está menos dispersa que la distribución Y

La figura siguiente presenta distribuciones X y Y para las que H_a es verdadera. La prueba Wilcoxon de suma de rangos no es apropiada en esta situación porque cuando H_a es verdadera como se ilustra, las Y tenderán a estar en los extremos máximos de la muestra combinada (lo que da por resultado rangos pequeños y grandes de Y), de modo que la suma de los rangos de X dará por resultado un valor W que no es ni grande ni pequeño.



Considere modificar el procedimiento para asignar rangos como sigue: después de ordenar la muestra combinada de $m + n$ observaciones, la observación más pequeña recibe un rango 1, la observación más grande recibe un rango 2, la segunda más pequeña recibe rango 3, la segunda más grande recibe rango 4, y así sucesivamente. Entonces, si H_a es verdadera como se ilustra, los valores X tenderán a estar en la parte media de la muestra y en esa forma reciben rangos grandes. Denote con W' la suma de los rangos X y considere rechazar H_0 a favor de H_a cuando $w' \geq c$. Cuando H_0 es verdadera, todo posible conjunto de rangos de X tiene la misma probabilidad, de modo que *W' tiene la misma distribución que W cuando H_0 es verdadera.* Entonces, c se puede seleccionar de la tabla A.14 del apéndice para dar una prueba de nivel α. La información siguiente se refiere al grosor muscular intermedio, para arteriolas de pulmones de niños que murieron de síndrome de muerte infantil repentina (x) (SIDS, por sus siglas en inglés) y un grupo de niños de control (y). Efectúe la prueba de H_0 contra H_a al nivel .05.

SIDS	4.0	4.4	4.8	4.9	
Control	3.7	4.1	4.3	5.1	5.6

Consulte el libro de Lehmann (en la bibliografía) para más información de esta prueba, llamada de *Siegel-Tukey.*

36. El procedimiento para ordenar que se describe en el ejercicio 35 es asimétrico en ocasiones, porque la observación más pequeña recibe un rango 1 mientras que la más grande recibe rango 2, y así sucesivamente. Suponga que la más pequeña y la más grande reciben rango 1, la segunda más pequeña y la segunda más grande reciben rango 2, y así sucesivamente, y sea W'' la suma de los rangos X. La distribución nula de W'' no es idéntica a la distribución nula de W, de modo que se hacen necesarias tablas diferentes. Considere el caso $m = 3$, $n = 4$. Haga una lista de los 35 posibles ordenamientos de los tres valores X entre las siete observaciones (por ejemplo, 1, 3, 7 o 4, 5, 6), asigne rangos en la forma descrita, calcule el valor de W'' para cada posibilidad, y a continuación tabule la distribución nula de W''. Para la prueba que rechaza si $w'' \geq c$, ¿qué valor de c prescribe aproximadamente una prueba de nivel .10? Ésta es la *prueba Ansari-Bradley*; para información adicional, vea el libro de Hollander y Wolfe en la bibliografía del capítulo.

Bibliografía

Hollander, Myles, and Douglas Wolfe, *Nonparametric Statistical Methods* (2a. ed.), Wiley, Nueva York, 1999. Una muy buena referencia sobre métodos de distribución libre con un excelente conjunto de tablas.

Lehmann, Erich, *Nonparametrics: Statistical Methods Based on Ranks*, Springer, Nueva York, 2006. Un excelente examen de los métodos de distribución libre más importantes, presentados con gran cantidad de ingeniosos comentarios.

16 Métodos de control de calidad

INTRODUCCIÓN

Las características de calidad de productos manufacturados han recibido considerable atención por parte de ingenieros de diseño y personal de producción, así como de quienes se ocupan de la administración financiera. Un artículo de fe por muchos años fue que los muy elevados niveles de calidad y bienestar económico eran metas comparables, pero recientemente se ha hecho cada vez más evidente que elevar los niveles de calidad puede llevar a reducir costos y a un mayor grado de satisfacción del consumidor, con lo cual aumenta la rentabilidad. Esto ha dado por resultado un renovado énfasis en técnicas estadísticas para diseñar calidad en productos y para identificar problemas de calidad en diversas etapas de producción y distribución.

Las gráficas de control tienen ahora un amplio uso como técnica de diagnóstico para el monitoreo de procesos de producción y de servicio, para identificar inestabilidad y circunstancias poco comunes. Después de una introducción a las ideas básicas en la sección 16.1, en las siguientes cuatro secciones se presentan diversas gráficas de control. La base para casi todas éstas se encuentra en el trabajo previo, referente a distribuciones de probabilidad de varias estadísticas, como por ejemplo la media muestral $\overline{X}$ y la proporción muestral $\hat{p} = X/n$.

Otra situación, que por lo general se encuentra en la industria, es una decisión tomada por un cliente en cuanto a si un lote de piezas ofrecidas por un proveedor es de calidad aceptable. En la última sección del capítulo se examinan brevemente algunos métodos de muestreo de aceptación para decidir, con base en datos muestrales, sobre la disposición de un lote.

Además de gráficas de control y planes de muestreo de aceptación, que primero se desarrollaron en las décadas de 1920 y 1930, expertos en estadística e ingenieros han introducido recientemente numerosos métodos estadísticos para identificar tipos y niveles de entradas de producción que aseguran una producción de alta calidad. Investigadores japoneses y en particular el ingeniero y estadístico

G. Taguchi y sus discípulos, han tenido gran influencia en este sentido; ahora hay una gran cantidad de información conocida como "métodos Taguchi". Las ideas de diseño experimental, y en particular experimentos de factores fraccionarios, son ingredientes clave. Todavía hay mucha controversia en la comunidad estadística en cuanto a cuáles diseños y métodos de análisis son más apropiados para el trabajo a ejecutar. Una crítica reciente está contenida en el artículo de George Box y otros citado en la bibliografía de este capítulo; el libro de Thomas Ryan citado aquí es también una buena fuente de información.

16.1 Comentarios generales sobre gráficas de control

Un mensaje esencial en todo este libro ha sido la capacidad de penetración de la variación que ocurre de manera natural, asociada con cualquier característica o atributo de individuos u objetos diferentes. En un contexto de manufacturas, no importa lo cuidadosamente que se hayan calibrado las máquinas, que se controlen los factores ambientales, se supervisen materiales y otros insumos, y se capaciten trabajadores, el diámetro variará de un tornillo a otro, algunas hojas de plástico serán más fuertes que otras, algunos fusibles serán defectuosos y otros no tendrán problemas, y así sucesivamente. Se podría considerar que estas variaciones aleatorias naturales son ruido de fondo incontrolable.

Son, no obstante, otras fuentes de variación las que pueden tener un impacto pernicioso en la calidad de piezas producidas por algún proceso. Esta variación puede ser atribuible a material contaminado, ajustes incorrectos de máquinas, desgaste poco común en herramientas, y otros semejantes. Estas fuentes de variación se han denominado *causas asignables* en la literatura de control de calidad. Las **gráficas de control** son un mecanismo para reconocer situaciones donde las causas asignables pueden estar afectando de manera adversa la calidad de un producto. Una vez que la gráfica indique una situación fuera de control, puede iniciarse una investigación para identificar causas y tomar medidas correctivas.

Un elemento básico de las gráficas de control es que las muestras se hayan seleccionado del proceso de interés en una secuencia de tiempos. Dependiendo del aspecto del proceso bajo investigación, se selecciona alguna estadística, por ejemplo la media muestral o la proporción muestral de los artículos defectuosos. El valor de esta estadística se calcula entonces para cada muestra en su turno. El resultado entonces es una gráfica de control tradicional al graficar estos valores calculados en el tiempo, como se ilustra en la figura 16.1.

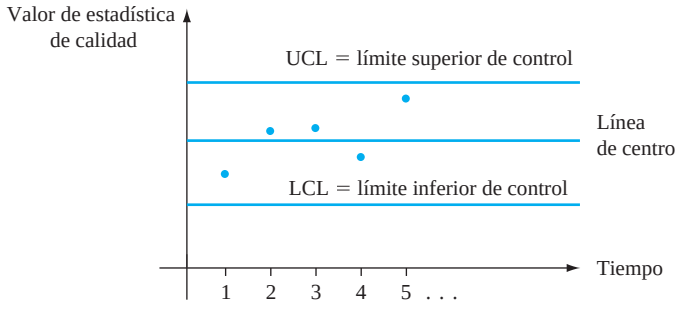


Figura 16.1 Prototipo de gráfica de control

Nótese que además de los puntos localizados, la gráfica tiene una línea de centro y dos límites de control. La base para la selección de una línea de centro es a veces un valor objetivo o especificación de diseño, por ejemplo, un valor deseado del diámetro de cojinetes. En otros casos, la altura de la línea de centro se estima a partir de la información. Si los puntos de la gráfica están todos entre los dos límites de control, se considera que el proceso está bajo control. Esto es, se piensa que el proceso está operando de un modo estable que refleja sólo una variación aleatoria natural. Una "señal" fuera de control se presenta siempre que los puntos graficados caigan fuera de los límites, lo cual se supone atribuible a alguna causa asignable y se inicia una búsqueda de esa causa. Los límites están diseñados de modo que un proceso bajo control genere muy pocas alarmas falsas, mientras que un proceso fuera de control rápidamente dé lugar a un punto fuera de los límites.

Hay una fuerte analogía entre la lógica de una gráfica de control y el trabajo previo de pruebas de hipótesis. La hipótesis nula aquí es que el proceso está bajo control. Cuando un proceso bajo control da un punto *fuera* de los límites de control (una señal fuera de control), ha ocurrido un error tipo I. Por otra parte, un error tipo II ocurre cuando un proceso fuera de control produce un punto *dentro* de límites de control. La apropiada selección de tamaño muestral y límites de control (estos últimos correspondientes a especificar una región de rechazo en la prueba de hipótesis) hará que las probabilidades de error asociadas sean adecuadamente pequeñas.

Hay que destacar que "bajo control" no es sinónimo de "satisface especificaciones o tolerancia de diseño". La magnitud de una variación natural puede ser tal que el porcentaje de piezas que no se apeguen a especificaciones sea mucho mayor de lo que se pueda tolerar. En tales casos, puede ser necesaria una reestructuración importante del proceso para mejorar la capacidad del proceso. Un proceso bajo control es simplemente aquel cuyo comportamiento con respecto a la variación es estable en el tiempo, sin mostrar indicaciones de causas extrañas poco comunes.

Ahora el software para elaborar gráficas de control es ampliamente utilizado. La revista *Quality Progress* contiene numerosos anuncios de paquetes de cómputo para el control estadístico de calidad. Además, SAS y la versión más reciente de Minitab, entre otros paquetes de uso general, tienen atractivas capacidades de control de calidad.

EJERCICIOS Sección 16.1 (1–5)

1. Una gráfica de control para el grosor de láminas de acero laminado está basada en un límite superior de control de .0520 pulgadas y un límite inferior de .0475 pulgadas. Los primeros diez valores del estadístico de control de calidad (en este caso $\overline{X}$, el grueso medio muestral de $n = 5$ láminas de muestra) son .0506, .0493, .0502, .0501, .0512, .0498, .0485, .0500, .0505 y .0483. Construya la parte inicial de la gráfica de control de calidad, y comente sobre su aspecto.

2. Consulte el ejercicio 1 y suponga que los diez valores más recientes de la estadística de control de calidad son .0493, .0485, .0490, .0503, .0492, .0486, .0495, .0494, .0493, y .0488. Construya la parte relevante de la correspondiente gráfica de control, y comente sobre su aspecto.

3. Suponga que se construye una gráfica de control de modo que la probabilidad de un punto que cae fuera de los límites de control, cuando el proceso está realmente bajo control, es .002. ¿Cuál es la probabilidad de que diez puntos sucesivos (con base en muestras seleccionadas de manera independiente) estén dentro de límites de control? ¿Cuál es la probabilidad de que 25 puntos sucesivos caigan todos dentro de límites de control? ¿Cuál es el número más pequeño de puntos sucesivos graficados, para el que la probabilidad de observar al menos uno fuera de límites de control exceda de .10?

4. Un corcho destinado a ser utilizado en una botella de vino se considera aceptable si su diámetro es de entre 2.9 cm y 3.1 cm (*por lo que el límite de especificación inferior* es LSL = 2.9 y *el límite de especificación superior* es USL = 3.1).
   a. Si el diámetro del corcho es una variable de una distribución normal con valor medio de 3.04 cm y desviación estándar de .02, ¿cuál es la probabilidad de que un corcho seleccionado al azar dé acuerdo a especificaciones?
   b. Si en cambio el valor medio es de 3.00 y la desviación estándar es de .05, ¿la probabilidad de que se ajuste a las especificaciones es más pequeña o más grande de lo que era en el inciso (a)?

5. Si una variable de proceso está distribuida normalmente, a la larga casi todos los valores observados deben estar entre $\mu - 3\sigma$ y $\mu + 3\sigma$, dando un proceso de propagación de 6σ.
   a. Con LSL y USL denotando los límites de especificación inferior y superior, se usa generalmente un *índice de capacidad de proceso* $C_p = (\text{USL} - \text{LSL})/6\sigma$. El valor $C_p = 1$ indica un proceso que es sólo marginalmente capaz de cumplir con las

especificaciones. Lo ideal es que C_p debe superar los 1.33 (un "muy buen" proceso). Calcule el valor de C_p para cada uno de los procesos de producción de corcho descritos en el ejercicio anterior y comente.

b. El índice C_p descrito en el inciso (a) no tiene en cuenta la localización del proceso. Una medida de capacidad que recurre a la media del proceso es

$$C_{pk} = \text{mín}\{(\text{USL} - \mu)/3\sigma, (\mu - \text{LSL})/3\sigma\}$$

Calcule el valor de C_{pk} para cada uno de los procesos de producción de corcho descritos en el ejercicio anterior y coméntelo. [*Nota*: en la práctica, μ y σ tienen que ser estimados a partir de datos del proceso, se muestra cómo hacer esto en la sección 16.2]

c. ¿Cómo se comparan C_p y C_{pk} y cuándo son iguales?

16.2 Gráficas de control para ubicación de proceso

Suponga que la característica de calidad de interés está asociada con una variable cuyos valores observados resultan de tomar mediciones. Por ejemplo, la característica podría ser la resistencia de un alambre eléctrico (ohms), el diámetro interno de juntas de expansión de caucho moldeado (cm), o la dureza de cierta aleación (unidades Brinell). Un uso importante de gráficas de control es ver si alguna medida de ubicación de la distribución de la variable permanece estable en el tiempo. La gráfica más popular para este fin es la $\bar{X}$.

La gráfica $\bar{X}$ basada en valores conocidos de parámetros

Debido a que hay incertidumbre acerca del valor de la variable para cualquier pieza o espécimen particular, con X se denota tal variable *aleatoria* (rv por sus siglas en inglés). Suponga que para un proceso bajo control, X tiene una distribución normal con valor medio μ y desviación estándar σ. Entonces, si $\bar{X}$ denota la media muestral para una muestra aleatoria de tamaño n seleccionada en un punto de tiempo particular, se sabe que

1. $E(\bar{X}) = \mu$

2. $\sigma_{\bar{X}} = \sigma/\sqrt{n}$

3. $\bar{X}$ tiene una distribución normal.

Se deduce que

$$P(\mu - 3\sigma_{\bar{X}} \le \bar{X} \le \mu + 3\sigma_{\bar{X}}) = P(-3.00 \le Z \le 3.00) = .9974$$

donde Z es una variable aleatoria normal estándar.* Es entonces muy probable que para un proceso bajo control, la media muestral caerá dentro de 3 desviaciones estándar para ($3\sigma_{\bar{X}}$) de la media μ del proceso.

Considere primero el caso en el que se conocen los valores de μ y σ. Suponga que en cada uno de los puntos de tiempo 1, 2, 3, . . . , existe una muestra aleatoria de tamaño n. Denote con $\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots$ los valores calculados de las medias muestrales correspondientes. Resulta una gráfica $\bar{X}$ de graficar estas $\bar{x}_i$ en el tiempo, es decir, graficar los puntos $(1, \bar{x}_1)$, $(2, \bar{x}_2)$, $(3, \bar{x}_3)$, y así sucesivamente, y luego trazar rectas horizontales de un lado a otro de la gráfica en

$$\text{LCL} = \text{límite inferior de control} = \mu - 3 \cdot \frac{\sigma}{\sqrt{n}}$$

$$\text{UCL} = \text{límite superior de control} = \mu + 3 \cdot \frac{\sigma}{\sqrt{n}}$$

*El uso de gráficas basadas en 3 límites de desviación estándar es tradicional, pero la tradición ciertamente no es inviolable.

Este trazado se llama con frecuencia gráfica 3 sigma. Cualquier punto fuera de los límites de control sugiere que el proceso pueda haber estado fuera de control en ese tiempo, de modo que debe iniciarse una búsqueda de las causas asignables.

Ejemplo 16.1 Una vez al día, se seleccionan aleatoriamente tres especímenes de aceite de motor del proceso de producción, y cada uno se analiza para determinar su viscosidad. La información siguiente (tabla 16.1) es para un periodo de 25 días. Una amplia experiencia con este proceso sugiere que cuando el proceso está bajo control, la viscosidad de un espécimen está normalmente distribuida con media 10.5 y desviación estándar .18. Entonces $\sigma_{\bar{X}} = \sigma/\sqrt{n} = .18/\sqrt{3} = .104$, de modo que los límites de control de 3 desviaciones estándar son

$$\text{LCL} = \mu - 3 \cdot \frac{\sigma}{\sqrt{n}} = 10.5 - 3(.104) = 10.188$$

$$\text{UCL} = \mu + 3 \cdot \frac{\sigma}{\sqrt{n}} = 10.5 + 3(.104) = 10.812$$

Tabla 16.1 Información de viscosidad para el ejemplo 16.1

Día	Observaciones de viscosidad			$\bar{x}$	s	Rango
1	10.37	10.19	10.36	10.307	.101	.18
2	10.48	10.24	10.58	10.433	.175	.34
3	10.77	10.22	10.54	10.510	.276	.55
4	10.47	10.26	10.31	10.347	.110	.21
5	10.84	10.75	10.53	10.707	.159	.31
6	10.48	10.53	10.50	10.503	.025	.05
7	10.41	10.52	10.46	10.463	.055	.11
8	10.40	10.38	10.69	10.490	.173	.31
9	10.33	10.35	10.49	10.390	.087	.16
10	10.73	10.45	10.30	10.493	.218	.43
11	10.41	10.68	10.25	10.447	.217	.43
12	10.00	10.60	10.71	10.437	.382	.71
13	10.37	10.50	10.34	10.403	.085	.16
14	10.47	10.60	10.75	10.607	.140	.28
15	10.46	10.46	10.56	10.493	.058	.10
16	10.44	10.68	10.32	10.480	.183	.36
17	10.65	10.42	10.26	10.443	.196	.39
18	10.73	10.72	10.83	10.760	.061	.11
19	10.39	10.75	10.27	10.470	.250	.48
20	10.59	10.23	10.35	10.390	.183	.36
21	10.47	10.67	10.64	10.593	.108	.20
22	10.40	10.55	10.38	10.443	.093	.17
23	10.24	10.71	10.27	10.407	.263	.47
24	10.37	10.69	10.40	10.487	.177	.32
25	10.46	10.35	10.37	10.393	.059	.11

Todos los puntos de la gráfica de control mostrados en la figura 16.2 están entre límites de control, lo que indica comportamiento estable de la media del proceso en el periodo (la desviación estándar y el rango para cada muestra se usarán en la subsección siguiente).

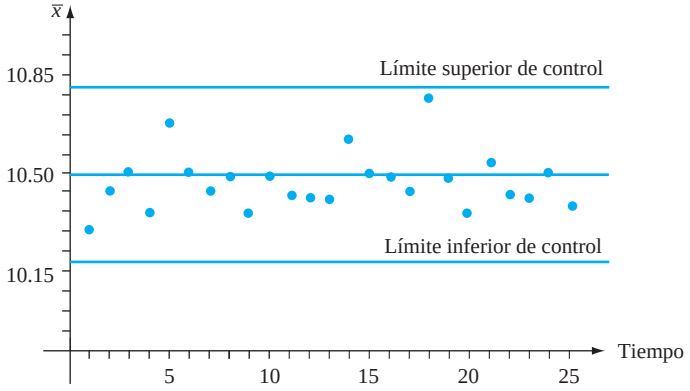


Figura 16.2 Gráfica $\bar{X}$ para la información de viscosidad del ejemplo 16.1

Gráficas $\bar{X}$ basadas en parámetros estimados

En la práctica ocurre con frecuencia que se desconocen valores de μ y σ, de modo que deben estimarse a partir de datos muestrales antes de determinar los límites de control. Esto es especialmente cierto cuando un proceso se somete primero a un análisis de control de calidad. Otra vez denote por n el número de observaciones de cada muestra y represente con k el número de muestras disponibles. Los valores típicos de n son 3, 4, 5, o 6, y se recomienda que k sea al menos 20. Se supone que las k muestras se reunieron durante un periodo cuando se pensaba que el proceso estaba bajo control. En breve se analizará más esta suposición.

Con $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$ denotando las k medias muestrales calculadas, la estimación común de μ es simplemente el promedio de estas medias:

$$\hat{\mu} = \bar{\bar{x}} = \frac{\sum_{i=1}^{k} \bar{x}_i}{k}$$

Se presentan dos métodos diferentes que por lo común se usan para estimar σ, uno de ellos basado en las k desviaciones muestrales estándar y el otro en los k rangos muestrales (recuerde que el rango muestral es la diferencia entre las observaciones muestrales más grandes y las más pequeñas). Antes de la generalizada disponibilidad de buenas calculadoras y software computarizado de estadística, la facilidad de hacer cálculos manualmente era una consideración de extrema importancia, de modo que predominaba el método del rango. No obstante, en el caso de una distribución normal de población, se sabe que el estimador insesgado de σ basado en S tiene una varianza más pequeña que el basado en el rango muestral. Los expertos en estadística dicen que el primer estimador es más *eficiente* que el último. La pérdida de eficiencia para el estimador es ligera cuando n es muy pequeña pero se hace importante cuando $n > 4$.

Recuerde que la desviación estándar muestral no es un estimador insesgado para σ. Cuando $X_1, \ldots, X_n$ es una muestra aleatoria de una distribución normal, se puede demostrar (vea el ejercicio 6.37) que

$$E(S) = a_n \cdot \sigma$$

donde

$$a_n = \frac{\sqrt{2}\Gamma(n/2)}{\sqrt{n-1}\Gamma[(n-1)/2]}$$

y $\Gamma(\cdot)$ denota la función gamma (vea la sección 4.4). A continuación aparece una tabulación de a_n para la n seleccionada:

n	3	4	5	6	7	8
a_n	.886	.921	.940	.952	.959	.965

Sea

$$\bar{S} = \frac{\sum_{i=1}^{k} S_i}{k}$$

donde $S_1, S_2, \ldots, S_k$ son las desviaciones estándar muestrales para las k muestras. Entonces

$$E(\bar{S}) = \frac{1}{K} E\left(\sum_{i=1}^{k} S_i\right) = \frac{1}{k}\sum_{i=1}^{k} E(S_i) = \frac{1}{k}\sum_{i=1}^{k} a_n \cdot \sigma = a_n \cdot \sigma$$

De este modo

$$E\left(\frac{\bar{S}}{a_n}\right) = \frac{1}{a_n} E(\bar{S}) = \frac{1}{a_n} \cdot a_n \cdot \sigma = \sigma$$

de modo que $\hat{\sigma} = \bar{S}/a_n$ es un estimador insesgado de σ.

Límites de control basados en las desviaciones muestrales estándar

$$\text{LCL} = \bar{\bar{x}} = 3 \cdot \frac{\bar{s}}{a_n\sqrt{n}}$$

$$\text{UCL} = \bar{\bar{x}} + 3 \cdot \frac{\bar{s}}{a_n\sqrt{n}}$$

donde

$$\bar{\bar{x}} = \frac{\sum_{i=1}^{k} \bar{x}_i}{k} \qquad \bar{s} = \frac{\sum_{i=1}^{k} s_i}{k}$$

Ejemplo 16.2 Si se consulta la información de viscosidad del ejemplo 16.1, se tenía $n = 3$ y $k = 25$. Los valores de $\bar{x}_i$ y $s_i (i = 1, \ldots, 25)$ aparecen en la tabla 16.1, de lo cual se deduce que $\bar{\bar{x}} = 261.896/25 = 10.476$ y $\bar{s} = 3.834/25 = .153$. Con $a_3 = .886$, se tiene

$$\text{LCL} = 10.476 - 3 \cdot \frac{.153}{.886\sqrt{3}} = 10.476 - .299 = 10.177$$

$$\text{UCL} = 10.476 + 3 \cdot \frac{.153}{.886\sqrt{3}} = 10.476 + .299 = 10.775$$

Estos límites difieren un poco de los límites previos basados en $\mu = 10.5$ y $\sigma = .18$ porque ahora $\hat{\mu} = 10.476$ y $\hat{\sigma} = \bar{s}/a_3 = .173$. Una inspección de la tabla 16.1 muestra que toda $\bar{x}_i$ está entre estos nuevos límites, de manera que de nueva cuenta no es evidente una situación fuera de control. ■

Para obtener una estimación de σ basada en el rango muestral, note que si $X_1, \ldots, X_n$ forman una muestra aleatoria de una distribución normal, entonces

$$\begin{aligned} R &= \text{rango}(X_1, \ldots, X_n) = \text{máx}(X_1, \ldots, X_n) - \text{mín}(X_1, \ldots, X_n) \\ &= \text{máx}(X_1 - \mu, \ldots, X_n - \mu) - \text{mín}(X_1 - \mu, \ldots, X_n - \mu) \\ &= \sigma\left\{\text{máx}\left(\frac{X_1 - \mu}{\sigma}, \cdots, \frac{X_n - \mu}{\sigma}\right) - \text{mín}\left(\frac{X_1 - \mu}{\sigma}, \cdots, \frac{X_n - \mu}{\sigma}\right)\right\} \\ &= \sigma \cdot \{\text{máx}(Z_1, \ldots, Z_n) - \text{mín}(Z_1, \ldots, Z_n)\} \end{aligned}$$

donde $Z_1, \ldots, Z_n$ son variables aleatorias normales estándar independientes. Entonces

$$\begin{aligned} E(R) &= \sigma \cdot E \text{ (rango de una muestra normal estándar)} \\ &= \sigma \cdot b_n \end{aligned}$$

de modo que R/b_n es un estimador insesgado de σ.

Ahora denote con $r_1, r_2, \ldots, r_k$ los rangos para las k muestras del conjunto de datos de control de calidad. El argumento que se acaba de dar implica que la estimación

$$\hat{\sigma} = \frac{\frac{1}{k}\sum_{i=1}^{k} r_i}{b_n} = \frac{\bar{r}}{b_n}$$

proviene de un estimador insesgado para σ. Los valores seleccionados de b_n aparecen en la tabla siguiente [su cálculo está basado en el uso de teoría estadística e integración numérica para determinar $E(\text{mín}(Z_1, \ldots, Z_n))$ y $E(\text{máx}(Z_1, \ldots, Z_n))$].

n	3	4	5	6	7	8
b_n	1.693	2.058	2.325	2.536	2.706	2.844

Límites de control basados en los rangos muestrales

$$\text{LCL} = \bar{\bar{x}} - 3 \cdot \frac{\bar{r}}{b_n\sqrt{n}}$$

$$\text{UCL} = \bar{\bar{x}} + 3 \cdot \frac{\bar{r}}{b_n\sqrt{n}}$$

donde $\bar{r} = \sum_{i=1}^{k} r_i/k$ y $r_1, \ldots, r_k$ son los k rangos muestrales individuales.

Ejemplo 16.3 (Continuación del ejemplo 16.2)

La tabla 16.1 da como resultado $\bar{r} = .292$, de modo que $\hat{\sigma} = .292/b_3 = .292/1.693 = .172$ y

$$\text{LCL} = 10.476 - 3 \cdot \frac{.292}{1.693\sqrt{3}} = 10.476 - .299 = 10.177$$

$$\text{UCL} = 10.476 + 3 \cdot \frac{.292}{1.693\sqrt{3}} = 10.476 + .299 = 10.775$$

Estos límites son idénticos a los basados en $\bar{s}$, y de nuevo toda $\bar{x}_i$ se encuentra entre límites. ■

Recálculo de límites de control

Se ha supuesto que la información muestral empleada para estimar μ y σ se obtuvo de un proceso bajo control, pero suponga que uno de los puntos de la gráfica de control resultante cae fuera de los límites de control. Entonces, si se puede hallar y verificar una causa

asignable para esta situación fuera de control, se recomienda que los nuevos límites de control se calculen después de borrar la muestra correspondiente del conjunto de datos. Del mismo modo, si más de un punto cae fuera de los límites originales, deben determinarse nuevos límites después de eliminar cualquier punto para el que una causa asignable pueda identificarse y resolverse. Incluso puede ocurrir que uno o más puntos caigan fuera de los nuevos límites, en cuyo caso el proceso de borrar y recalcular debe repetirse.

Características de operación de gráficas de control

En términos generales, una gráfica de control será efectiva si da muy pocas señales fuera de control cuando el proceso está bajo control, pero muestra un punto fuera de los límites de control casi tan pronto como el proceso se sale de control. Una evaluación de la efectividad de una gráfica está basada en la noción de "probabilidades de error". Suponga que la variable de interés está distribuida normalmente con σ conocida (el mismo valor para un proceso bajo control o fuera de control). Además, considere una gráfica de 3 sigmas basada en el valor objetivo μ_0, con $\mu = \mu_0$ cuando el proceso está bajo control. Una probabilidad de error es

$$\begin{aligned}\alpha &= P(\text{una sola muestra da un punto fuera de límites de control cuando } \mu = \mu_0)\\ &= P(\bar{X} > \mu_0 + 3\sigma/\sqrt{n} \text{ o } \bar{X} < \mu_0 - 3\sigma/\sqrt{n} \text{ cuando } \mu = \mu_0)\\ &= P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > 3 \text{ o } \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} < -3 \text{ cuando } \mu = \mu_0\right)\end{aligned}$$

La variable estandarizada $Z = (\bar{X} - \mu_0)/(\sigma/\sqrt{n})$ tiene una distribución normal estándar cuando $\mu = \mu_0$, de modo que

$$\alpha = P(Z > 3 \text{ o } Z < -3) = \Phi(-3.00) + 1 - \Phi(3.00) = .0026$$

Si se ha usado 3.09 en lugar de 3 para determinar los límites de control (esto se acostumbra en la Gran Bretaña), entonces

$$\alpha = P(Z > 3.09 \text{ o } Z < -3.09) = .0020$$

El uso de límites de 3 sigmas hace muy poco probable que resulte una señal fuera de control de un proceso bajo control.

Ahora suponga que el proceso está fuera de control porque μ ha cambiado a $\mu + \Delta\sigma$ (Δ puede ser positiva o negativa); Δ es el número de desviaciones estándar en los que μ ha cambiado. Una segunda probabilidad de error es

$$\begin{aligned}\beta &= P\left(\begin{array}{l}\text{una sola muestra da un punto dentro de}\\ \text{límites de control cuando } \mu = \mu_0 + \Delta\sigma\end{array}\right)\\ &= P(\mu_0 - 3\sigma/\sqrt{n} < \bar{X} < \mu_0 + 3\sigma/\sqrt{n} \text{ cuando } \mu = \mu_0 + \Delta\sigma)\end{aligned}$$

Ahora se estandariza al restar primero $\mu_0 + \Delta\sigma$ de cada término que está dentro de los paréntesis y a continuación dividir entre $\sigma/\sqrt{n}$:

$$\begin{aligned}\beta &= P(-3 - \sqrt{n}\Delta < \text{variable aleatoria normal estándar} < 3 - \sqrt{n}\Delta)\\ &= \Phi(3 - \sqrt{n}\Delta) - \Phi(-3 - \sqrt{n}\Delta)\end{aligned}$$

Esta probabilidad de error depende de Δ, que determina el tamaño del cambio, y del tamaño muestral n. En particular, para Δ fija, β disminuirá cuando n aumente (cuanto más grande sea el tamaño muestral, es más probable que resulte una señal fuera de control), y para n fija, β disminuye cuando $|\Delta|$ aumenta (cuanto mayor sea la magnitud del cambio, es más probable que resulte una señal fuera de control). La tabla siguiente da β para valores seleccionados de Δ cuando $n = 4$.

Δ	.25	.50	.75	1.00	1.50	2.00	2.50	3.00
β ***cuando*** $n = 4$	.9936	.9772	.9332	.8413	.5000	.1587	.0668	.0013

Es evidente que es muy poco probable que un pequeño cambio pase sin ser detectado en una muestra individual.

Si 3 es sustituido por 3.09 en los límites de control, entonces α disminuye de .0026 a .002, pero para cualesquiera n y σ fijas, β aumentará. Esto es sólo una manifestación de la relación inversa entre los dos tipos de probabilidades de error al probar hipótesis. Por ejemplo, cambiar 3 a 2.5 aumentará α y disminuirá β.

Las probabilidades de error examinadas hasta aquí se calculan bajo la suposición de que la variable de interés está distribuida de manera normal. Si la distribución es sólo ligeramente no normal, el efecto del teorema de límite central implica que $\overline{X}$ tendrá de un modo aproximado una distribución normal incluso cuando n sea pequeña, en cuyo caso las probabilidades de error indicadas serán casi correctas. Éste, por supuesto, ya no es el caso cuando la distribución de la variable se desvía considerablemente de la normalidad.

Una segunda evaluación del uso comprende la dimensión esperada o promedio de lote, necesaria para observar una señal fuera de control. Cuando el proceso está bajo control, se debería esperar observar muchas muestras antes de ver una cuya $\bar{x}$ se encuentre fuera de los límites de control. Por otra parte, si un proceso se sale de control, el número esperado de muestras necesario para detectar esto debe ser pequeño.

Denote con p la probabilidad de que una sola muestra dé un valor $\bar{x}$ fuera de los límites de control, es decir,

$$p = P(\overline{X} < \mu_0 - 3\sigma/\sqrt{n} \quad \text{o} \quad \overline{X} > \mu_0 + 3\sigma/\sqrt{n})$$

Considere primero un proceso bajo control, de modo que $\overline{X}_1, \overline{X}_2, \overline{X}_3, \ldots$ están todas normalmente distribuidas con valor medio μ_0 y desviación estándar $\sigma/\sqrt{n}$. Defina una variable aleatoria Y por

$$Y = \text{la primera } i \text{ para la que } \overline{X}_i \text{ cae fuera de límites de control}$$

Si se considera cada número muestral como un intento y una muestra fuera de control como un éxito, entonces Y es el número de intentos (independientes) necesarios para observar un éxito. Esta Y tiene una distribución geométrica, como se demostró en el ejemplo 3.18 que $E(Y) = 1/p$. El acrónimo ARL (por sus siglas en inglés de *magnitud promedio de lote*) se usa con frecuencia en lugar de $E(Y)$. Como $p = \alpha$ para un proceso bajo control, se tiene

$$\text{ARL} = E(Y) = \frac{1}{p} = \frac{1}{\alpha} = \frac{1}{.0026} = 384.62$$

Al sustituir 3 en los límites de control con 3.09, resulta ARL = 1/.002 = 500.

Ahora suponga que, en un punto particular de tiempo, la media del proceso cambia a $\mu = \mu_0 + \Delta\sigma$. Si se define Y como la primera i subsiguiente al cambio para el que una muestra genera una señal fuera de control, de nuevo es cierto que ARL $= E/Y = 1/p$, pero ahora $p = 1 - \beta$. La tabla siguiente da las ARL seleccionadas para una gráfica de 3 sigmas cuando $n = 4$. Estos resultados nuevamente demuestran la efectividad de la gráfica para detectar cambios grandes pero también su incapacidad de identificar con rapidez cambios pequeños. Cuando el muestreo se hace con poca frecuencia, es probable que muchas piezas se produzcan antes que un pequeño cambio en μ sea detectado. Los procedimientos CUSUM examinados en la sección 16.5 se crearon para manejar esta deficiencia.

Δ	.25	.50	.75	1.00	1.50	2.00	2.50	3.00
ARL cuando $n = 4$	156.25	43.86	14.97	6.30	2.00	1.19	1.07	1.0013

Reglas suplementarias para gráficas $\bar{X}$

La incapacidad de gráficas $\bar{X}$ con límites de 3 sigmas, para detectar con rapidez cambios pequeños en la media de un proceso, ha impulsado a investigadores a crear procesos que den un mejor comportamiento en este aspecto. Un método consiste en la introducción de condiciones adicionales que hacen que se genere una señal fuera de control. Las siguientes condiciones fueron recomendadas por Western Electric (entonces subsidiaria de AT&T). Una intervención para tomar medidas correctivas es apropiada siempre que se satisfaga una de estas condiciones:

1. Dos de cada tres puntos sucesivos caigan fuera de límites de 2 sigmas en el mismo lado de la línea de centro.
2. Cuatro de cada cinco puntos sucesivos caigan fuera de límites de 1 sigma en el mismo lado de la línea de centro.
3. Ocho puntos sucesivos caigan fuera en el mismo lado de la línea de centro.

Debe consultarse un texto sobre control de calidad para ver una discusión de éstas y otras reglas suplementarias.

Gráficas de control robustas

La presencia de valores aislados en la información de la muestra tiende a reducir la sensibilidad de procedimientos de gráficas de control cuando deban estimarse parámetros. Esto es porque los límites de control se mueven hacia fuera desde la línea de centro, lo cual hace más difícil la identificación de puntos poco comunes. *No* es deseable que la estadística cuyos valores se grafiquen sea resistente a valores aislados, porque esto ocultaría cualquier señal fuera de control. Por ejemplo, graficar medias muestrales sería menos eficaz que graficar $\bar{x}_1, \bar{x}_2, \ldots$ como se hace en una gráfica $\bar{X}$.

El artículo "Robust Control Charts" de David M. Rocke (*Technometrics,* 1989: 173–184) presenta un estudio de procedimientos para los que los límites de control están basados en estadísticas resistentes a los efectos de valores aislados. Rocke recomienda límites de control calculados del *rango intercuartil* (IQR), que es muy semejante a la cuarta dispersión introducida en el capítulo 1. En particular,

$$\text{IQR} = \begin{cases} (\text{2a. } x_i \text{ más grande}) - (2^{\text{a}}\ x_i \text{ más pequeña}) & n = 4, 5, 6, 7 \\ (\text{3a. } x_i \text{ más grande}) - (3^{\text{a}}\ x_i \text{ más pequeña}) & n = 8, 9, 10, 11 \end{cases}$$

Para una muestra aleatoria de una distribución normal, $E(\text{IQR}) = k_n\sigma$; los valores de k_n se dan en la tabla siguiente:

n	4	5	6	7	8
k_n	.596	.990	1.282	1.512	.942

Los límites de control sugeridos son

$$\text{LCL} = \bar{\bar{x}} - 3 \cdot \frac{\text{IQR}}{k_n\sqrt{n}} \qquad \text{UCL} = \bar{\bar{x}} + 3 \cdot \frac{\text{IQR}}{k_n\sqrt{n}}$$

Los valores de $\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots$ se grafican. Las simulaciones publicadas en el artículo indicaron que el uso de la gráfica con estos límites es superior al de la gráfica $\bar{X}$ tradicional.

EJERCICIOS Sección 16.2 (6–15)

6. En el caso de μ y σ conocidas, ¿qué límites de control son necesarios para que sea .005 la probabilidad de que un solo punto esté fuera de límites para un proceso bajo control?

7. Considere una gráfica de control de 3 sigmas con línea de centro en μ_0 y basada en $n = 5$. Suponiendo normalidad, calcule la probabilidad de que un solo punto caiga fuera de límites de control cuando la media real del proceso sea
a. $\mu_0 + .5\sigma$
b. $\mu_0 - \sigma$
c. $\mu_0 + 2\sigma$

8. La tabla de abajo da información sobre el contenido de humedad para especímenes de cierto tipo de tela. Determine límites de control para una gráfica con línea de centro a una altura de 13.00 con base en $\sigma = .600$, construya la gráfica de control y comente sobre su aspecto.

9. Consulte la información dada en el ejercicio 8 y construya una gráfica de control con una línea de centro estimada y límites basados en el uso de desviaciones muestrales estándar para estimar σ. ¿Hay alguna evidencia de que el proceso esté fuera de control?

10. Consulte los ejercicios 8 y 9, y ahora utilice límites de control con base en el uso de rangos muestrales para estimar σ. ¿El proceso parece estar bajo control?

11. La tabla siguiente da medias muestrales y desviaciones estándar, cada una basada en $n = 6$ observaciones del índice de refracción de cable de fibra óptica. Construya una gráfica de control, y comente sobre su aspecto. [*Sugerencia:* $\sum \bar{x}_i = 2317.07$ y $\sum s_i = 30.34$.]

Día	$\bar{x}$	s	Día	$\bar{x}$	s
1	95.47	1.30	13	97.02	1.28
2	97.38	.88	14	95.55	1.14
3	96.85	1.43	15	96.29	1.37
4	96.64	1.59	16	96.80	1.40
5	96.87	1.52	17	96.01	1.58
6	96.52	1.27	18	95.39	.98
7	96.08	1.16	19	96.58	1.21
8	96.48	.79	20	96.43	.75
9	96.63	1.48	21	97.06	1.34
10	96.50	.80	22	98.34	1.60
11	97.22	1.42	23	96.42	1.22
12	96.55	1.65	24	95.99	1.18

12. Consulte el ejercicio 11. Se encontró una causa asignable para el inusualmente alto índice de refracción promedio muestral el día 22. Recalcule límites de control después de borrar la información de este día. ¿Qué concluye el lector?

13. Considere la gráfica de control basada en límites de control $\mu_0 \pm 2.81\,\sigma/\sqrt{n}$.
a. ¿Cuál es la ARL cuando el proceso está bajo control?

Datos para el ejercicio 8

Muestra núm.	Observaciones de contenido de humedad					$\bar{x}$	s	Rango
1	12.2	12.1	13.3	13.0	13.0	12.72	.536	1.2
2	12.4	13.3	12.8	12.6	12.9	12.80	.339	.9
3	12.9	12.7	14.2	12.5	12.9	13.04	.669	1.7
4	13.2	13.0	13.0	12.6	13.9	13.14	.477	1.3
5	12.8	12.3	12.2	13.3	12.0	12.52	.526	1.3
6	13.9	13.4	13.1	12.4	13.2	13.20	.543	1.5
7	12.2	14.4	12.4	12.4	12.5	12.78	.912	2.2
8	12.6	12.8	13.5	13.9	13.1	13.18	.526	1.3
9	14.6	13.4	12.2	13.7	12.5	13.28	.963	2.4
10	12.8	12.3	12.6	13.2	12.8	12.74	.329	.9
11	12.6	13.1	12.7	13.2	12.3	12.78	.370	.9
12	13.5	12.3	12.8	13.1	12.9	12.92	.438	1.2
13	13.4	13.3	12.0	12.9	13.1	12.94	.559	1.4
14	13.5	12.4	13.0	13.6	13.4	13.18	.492	1.2
15	12.3	12.8	13.0	12.8	13.5	12.88	.432	1.2
16	12.6	13.4	12.1	13.2	13.3	12.92	.554	1.3
17	12.1	12.7	13.4	13.0	13.9	13.02	.683	1.8
18	13.0	12.8	13.0	13.3	13.1	13.04	.182	.5
19	12.4	13.2	13.0	14.0	13.1	13.14	.573	1.6
20	12.7	12.4	12.4	13.9	12.8	12.84	.619	1.5
21	12.6	12.8	12.7	13.4	13.0	12.90	.316	.8
22	12.7	13.4	12.1	13.2	13.3	12.94	.541	1.3

b. ¿Cuál es la ARL cuando $n = 4$ y la media del proceso se ha cambiado a $\mu = \mu_0 + \sigma$?

c. ¿Cómo se comparan los valores de los incisos (a) y (b) en función de los valores correspondientes para una gráfica de 3 sigmas?

14. Aplique las reglas suplementarias, sugeridas en el texto, a la información del ejercicio 8. ¿Hay alguna señal fuera de control?

15. Calcule límites de control para la información del ejercicio 8, usando el procedimiento robusto presentado en esta sección.

16.3 Gráficas de control para variación de proceso

Las gráficas de control estudiadas en la sección previa se diseñaron para controlar la ubicación (o bien, lo que es lo mismo, la tendencia central) de un proceso, con particular atención a la media como medida de ubicación. Es igualmente importante asegurar que un proceso está bajo control con respecto a variaciones. De hecho, casi todos los expertos recomiendan que se establezca un control sobre variaciones *antes* de construir una gráfica $\bar{X}$, o cualquier otra gráfica para controlar la ubicación. En esta sección, se consideran gráficas para variaciones con base en la desviación muestral estándar S y también gráficas basadas en el rango muestral R. Generalmente se prefieren las primeras porque la desviación estándar da una evaluación más eficiente de variación que el rango, pero las gráficas R se usaron primero y prevalece la tradición.

La gráfica S

De nuevo se supone que se dispone de k muestras seleccionadas de manera independiente, cada una formada por n observaciones en una variable distribuida normalmente. Denote las desviaciones muestrales estándar por $s_1, s_2, \ldots, s_k$, con $\bar{s} = \sum s_i/k$. Los valores s_1, s_2, s_3, . . . se grafican en secuencia en una gráfica S. La línea de centro de la gráfica estará a una altura $\bar{s}$, y los límites de 3 sigmas necesitan determinar $3\sigma_S$ (al igual que los límites de 3 sigmas de una gráfica $\bar{X}$ requirieron $3\sigma_{\bar{X}} = 3\sigma/\sqrt{n}$, con σ entonces estimada a partir de la información).

Recuerde que para cualquier variable aleatoria Y, $V(Y) = E(Y^2) - [E(Y)]^2$, y que una varianza muestral S^2 es un estimador insesgado de σ^2, es decir, $E(S^2) = \sigma^2$. Entonces,

$$V(S) = E(S^2) - [E(S)]^2 = \sigma^2 - (a_n\sigma)^2 = \sigma^2(1 - a_n^2)$$

donde los valores de a_n para $n = 3, \ldots, 8$ se tabularon en la sección previa. La desviación estándar de S es entonces

$$\sigma_S = \sqrt{V(S)} = \sigma\sqrt{1 - a_n^2}$$

Es natural estimar σ usando $s_1, \ldots, s_k$ como se hizo en la sección previa, es decir, $\hat{\sigma} = \bar{s}/a_n$. Sustituyendo $\hat{\sigma}$ por σ en la expresión para σ_S resulta la cantidad empleada para calcular límites de 3 sigmas.

Los límites de control de 3 sigmas para una gráfica de control S son

$$\text{LCL} = \bar{s} - 3\bar{s}\sqrt{1 - a_n^2}/a_n$$

$$\text{UCL} = \bar{s} - 3\bar{s}\sqrt{1 - a_n^2}/a_n$$

La expresión para LCL será negativa si $n \leq 5$, en cuyo caso se acostumbra usar LCL = 0,

Ejemplo 16.4 La tabla 16.2 muestra observaciones sobre la resistencia al esfuerzo en láminas de plástico (la fuerza, en libras por pulgada cuadrada, necesaria para agrietar una lámina). Hay $k = 22$ muestras, obtenidas en puntos de tiempo igualmente espaciados y $n = 4$ observaciones en cada muestra. Con facilidad se verifica que

$\sum s_i = 51.10$ y $\bar{s} = 2.32$, de modo que el centro de la gráfica S estará en 2.32 (aun cuando $n = 4$, el límite inferior de control (LCL) $= 0$ y la línea de centro no estará a la mitad entre los límites de control). De la sección previa, $a_4 = .921$, de donde el límite superior de control (UCL) es

$$\text{UCL} = 2.32 + 3(2.32)(\sqrt{1 - (.921)^2})/.921 = 5.26$$

Tabla 16.2 Datos sobre resistencia al esfuerzo para el ejemplo 16.4

Muestra núm.	Observaciones				Desviación estándar	Rango
1	29.7	29.0	28.8	30.2	.64	1.4
2	32.2	29.3	32.2	32.9	1.60	3.6
3	35.9	29.1	32.1	31.3	2.83	6.8
4	28.8	27.2	28.5	35.7	3.83	8.5
5	30.9	32.6	28.3	28.3	2.11	4.3
6	30.6	34.3	34.8	26.3	3.94	8.5
7	32.3	27.7	30.9	27.8	2.30	4.6
8	32.0	27.9	31.0	30.8	1.76	4.1
9	24.2	27.5	28.5	31.1	2.85	6.9
10	33.7	24.4	34.3	31.0	4.53	9.9
11	35.3	33.2	31.4	28.0	3.09	7.3
12	28.1	34.0	31.0	30.8	2.41	5.9
13	28.7	28.9	25.8	29.7	1.71	3.9
14	29.0	33.0	30.2	30.1	1.71	4.0
15	33.5	32.6	33.6	29.2	2.07	4.4
16	26.9	27.3	32.1	28.5	2.37	5.2
17	30.4	29.6	31.0	33.8	1.83	4.2
18	29.0	28.9	31.8	26.7	2.09	5.1
19	33.8	30.9	31.7	28.2	2.32	5.6
20	29.7	27.9	29.1	30.1	.96	2.2
21	27.9	27.7	30.2	32.9	2.43	5.2
22	30.0	31.4	27.7	28.1	1.72	3.7

La gráfica de control resultante se ilustra en la figura 16.3. Todos los puntos graficados están bien dentro de los límites de control, lo que sugiere un comportamiento estable del proceso con respecto a variaciones.

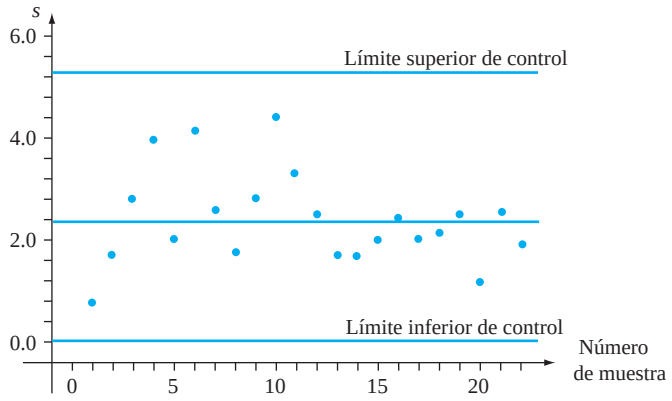


Figura 16.3 Gráfica S para datos de resistencia al esfuerzo del ejemplo 16.4 ■

La gráfica *R*

Denote con $r_1, r_2, \ldots, r_k$ los k rangos muestrales y $\bar{r} = \sum r_i/k$. La línea de centro de una gráfica R estará a una altura $\bar{r}$. La determinación de límites de control requiere σ_R, donde

R denota el rango (antes de hacer observaciones, como una variable aleatoria) de una muestra aleatoria de tamaño n de una distribución normal, con valor medio μ y desviación estándar σ. Como

$$\begin{aligned} R &= \text{máx}(X_1, \ldots, X_n) - \text{mín}(X_1, \ldots, X_n) \\ &= \sigma\{\text{máx}(Z_1, \ldots, Z_n) - \text{mín}(Z_1, \ldots, Z_n)\} \end{aligned}$$

donde $Z_i = (X_i - \mu)/\sigma$, y las Z_i son variables aleatorias normales estándar, se deduce que

$$\begin{aligned} \sigma_R &= \sigma \cdot \left(\begin{array}{l} \text{desviación estándar del rango de una muestra aleatoria} \\ \text{de tamaño } n \text{ de una distribución normal estándar} \end{array} \right) \\ &= \sigma \cdot c_n \end{aligned}$$

Los valores de c_n para $n = 3, \ldots, 8$ aparecen en la tabla siguiente.

n	3	4	5	6	7	8
c_n	.888	.880	.864	.848	.833	.820

Se acostumbra estimar σ por medio de $\hat{\sigma} = \bar{r}/b_n$ como se vio en la sección previa. Esto da $\hat{\sigma}_R = c_n\bar{r}/b_n$ como la desviación estándar estimada de R.

Los límites de 3 sigmas para una gráfica R son

$$\text{LCL} = \bar{r} - 3c_n\bar{r}/b_n$$
$$\text{UCL} = \bar{r} + 3c_n\bar{r}/b_n$$

La expresión para LCL será negativa si $n \leq 6$, en cuyo caso debería usarse LCL = 0.

Ejemplo 16.5 En la ingeniería de tejidos, las células se siembran sobre un andamio que a continuación, guía el crecimiento de nuevas células. El artículo "On the Process Capability of the Solid Free-Form Fabrication: A Case Study of Scaffold Moulds for Tissue Engineering" (*J. of Engr. in Med.*, 2008: 377–392) utiliza diversos métodos de control de calidad para estudiar un método de producir tales andamios. Una característica inusual es que en lugar de los subgrupos observados al paso del tiempo, cada subgrupo es el resultado de la dimensión (micras) de un diseño diferente. La tabla 16.3 contiene datos de la tabla 2 del artículo citado en la desviación del objetivo en la orientación perpendicular (estas desviaciones son de hecho todas positivas, los destellos impresos presentan dimensiones más grandes que los diseñados).

Tabla 16.3 Datos de desviación del objetivo para el ejemplo 16.5

dim dis				media	rango	desv est
200	12	17	6	11.7	11	5.51
250	6	9	17	10.7	11	5.69
300	5	9	15	9.7	10	5.03
350	19	6	11	12.0	13	6.56
400	9	14	9	10.7	5	2.89
450	9	15	8	10.7	7	3.79
500	8	11	12	10.3	4	2.08
550	4	14	11	9.7	10	5.13
600	11	14	7	10.7	7	3.51

(*continúa*)

Tabla 16.3 Datos de desviación del objetivo para el ejemplo 16.5 (continuación)

dim dis				media	rango	desv est
650	13	9	9	10.3	4	2.31
700	10	14	8	10.7	6	3.06
750	8	9	4	7.0	5	2.65
800	14	7	9	10.0	7	3.61
850	7	9	12	9.3	5	2.52
900	14	5	8	9.0	9	4.58
950	10	12	10	10.7	2	1.15
1000	7	11	15	11.0	8	4.00

La tabla 16.3 da $\sum r_i = 124$ de donde $\bar{r} = 7.29$. Como $n = 3$, el límite inferior de control $= 0$. Con $b_3 = 1.693$ y $c_3 = .888$

$$\text{UCL} = 7.29 + 3 \cdot (.888)(7.29)/1.693 = 18.76$$

La gráfica R y una gráfica $\overline{X}$ de Minitab aparecen en la figura 16.4 (el artículo citado también incluye estas gráficas). Todos los puntos están dentro de los límites de control apropiados, indicando un proceso controlado para las dos ubicaciones y variaciones.

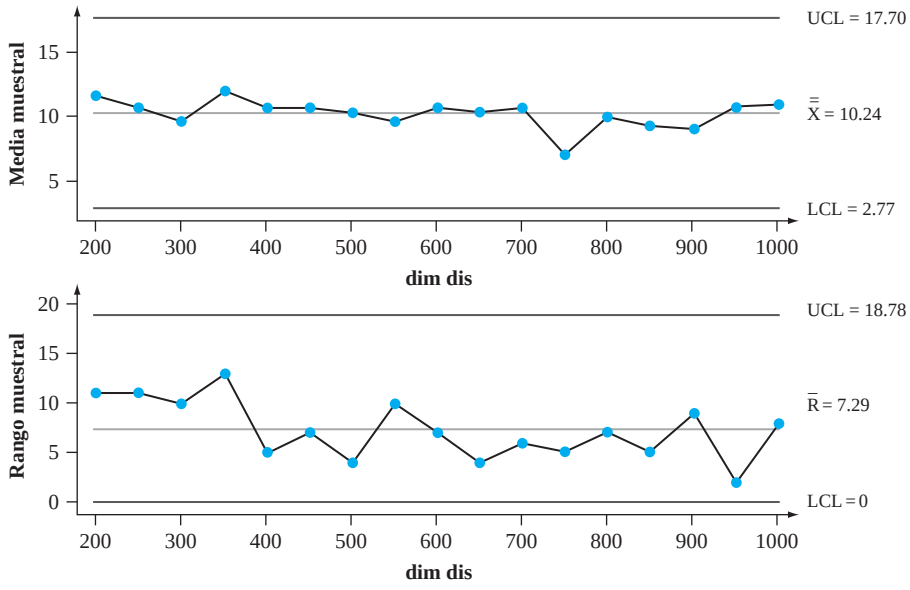


Figura 16.4 Gráficas de control para datos de desviación del objetivo para el ejemplo 16.5 ■

Gráficas basadas en límites de probabilidad

Considere una gráfica $\overline{X}$ basada en el valor μ_0 e incógnita σ bajo control (objetivo). Cuando la variable de interés está normalmente distribuida y el proceso está bajo control,

$$P(\overline{X}_i > \mu_0 + 3\sigma/\sqrt{n}) = .0013 = P(\overline{X}_i < \mu_0 - 3\sigma/\sqrt{n})$$

Esto es, la probabilidad de que un punto en la gráfica caiga arriba del límite superior de control es .0013, como es la probabilidad de que el punto caiga abajo del límite inferior de control (si se usa 3.09 en lugar de 3 resulta .001 para cada probabilidad). Cuando los

límites de control se basan en estimaciones de μ y σ, estas probabilidades serán aproximadamente correctas siempre que n no sea demasiado pequeña y k sea al menos 20.

En contraste, *no* es el caso para una gráfica S de 3 sigmas que $P(S_i > \text{UCL}) = P(S_i < \text{LCL}) = .0013$, ni es verdadero para una gráfica R de 3 sigmas que $P(R_i > \text{UCL}) = P(R_i < \text{LCL}) = .0013$. Esto es porque ni la desviación estándar muestral S ni el rango muestral R tienen una distribución normal aun cuando la distribución poblacional sea normal. En cambio, tanto S como R tienen distribuciones sesgadas. Lo mejor que se puede decir para gráficas S y R de 3 sigmas es que es bastante improbable que un proceso bajo control dé un punto, en cualquier tiempo particular, que esté fuera de los límites de control. Algunos autores han estado a favor del uso de límites de control para los cuales la "probabilidad de rebase" para cada límite es aproximadamente .001. El libro *Statistical Methods for Quality Improvement* (vea la bibliografía del capítulo) contiene más información sobre este tema.

EJERCICIOS Sección 16.3 (16–20)

16. Un fabricante de tiza sin polvo instituyó un programa de control de calidad para vigilar la densidad de la tiza. Las desviaciones muestrales estándar de densidades de 24 subgrupos diferentes, cada uno formado por $n = 8$ especímenes de tiza, fueron como sigue:

.204	.315	.096	.184	.230	.212	.322	.287
.145	.211	.053	.145	.272	.351	.159	.214
.388	.187	.150	.229	.276	.118	.091	.056

Calcule límites para una gráfica S, construya la gráfica, y verifique que no haya puntos fuera de control. Si hay un punto fuera de control, bórrelo y repita el proceso.

17. Una vez por hora, de una línea de ensamble se seleccionan subgrupos de unidades de fuentes de alimentación para determinar la salida de alto voltaje de cada unidad.

a. Suponga que la suma de las variaciones muestrales resultantes para 30 subgrupos, cada uno formado por cuatro unidades, es 85.2. Calcule los límites de control para una gráfica R.

b. Repita el inciso (a) si cada subgrupo está formado por ocho unidades y la suma es 106.2.

18. Los siguientes datos sobre la desviación de la meta en la orientación paralela es tomado de la tabla 1 del artículo citado en el ejemplo 16.5. A veces, una transformación de los datos es conveniente, ya sea por la no normalidad o porque la variación del subgrupo cambia sistemáticamente con la media del subgrupo. Los autores del artículo citado sugieren una transformación de raíz cuadrada de estos datos (*la familia de transformaciones de Box-Cox* es $y = x^{\lambda}$, así que aquí $\lambda = .5$; Minitab identificará el mejor valor de λ). Transforme los datos como se sugiere, calcule los límites de control para $\bar{X}$, R y gráfica S, y compruebe la presencia de cualquier señal fuera de control.

dim dis	observaciones		
200	14	31	12
250	22	13	9
300	12	22	16
350	11	28	1
400	15	12	36
450	6	31	14
500	13	24	9
550	21	18	16
600	6	16	20
650	8	17	23
700	3	26	17
750	17	12	22
800	41	17	3
850	18	11	21
900	9	15	22
950	25	4	17
1000	8	23	15

19. Calcule límites de control para una gráfica S desde la información de índice de refracción del ejercicio 11. ¿El proceso parece estar bajo control con respecto a la variabilidad? ¿Por qué sí o por qué no?

20. Cuando S^2 es la varianza muestral de una muestra aleatoria normal, $(n-1)S^2/\sigma^2$ tiene una distribución chi cuadrada con $n-1$ grados de libertad, de modo que

$$P\left(\chi^2_{.999,n-1} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{.001,n-1}\right) = .998$$

de la que

$$P\left(\frac{\sigma^2\chi^2_{.999,n-1}}{n-1} < S^2 < \frac{\sigma^2\chi^2_{.001,n-1}}{n-1}\right) = .998$$

Esto sugiere que una gráfica alternativa, para controlar la variación del proceso, consiste en graficar las varianzas muestrales y usar los límites de control

$$\text{LCL} = \bar{s^2}\chi^2_{.999,n-1}/(n-1)$$
$$\text{UCL} = \bar{s^2}\chi^2_{.001,n-1}/(n-1)$$

Construya la gráfica correspondiente para los datos del ejercicio 11. [*Sugerencia:* los valores críticos de chi cuadrada de colas inferior y superior para 5 grados de libertad son .210 y 20.515, respectivamente.]

16.4 Gráficas de control para atributos

El término *datos de atributos* se usa en literatura de control de calidad para describir dos situaciones:

1. Cada pieza producida es defectuosa o no defectuosa (se apega a especificaciones o no se apega).
2. Una sola pieza puede tener uno o más defectos, y el número de defectos se determina.

En el primer caso, una gráfica de control está basada en la distribución binomial; en el último, la distribución Poisson es la base para una gráfica.

La gráfica *p* para fracción defectuosa

Suponga que cuando un proceso está bajo control, la probabilidad de que cualquier pieza sea defectuosa es p (o bien, lo que es lo mismo, p es la proporción a largo plazo de piezas defectuosas para un proceso bajo control) y que diferentes piezas son independientes una de otra con respecto a sus condiciones. Considere una muestra de n piezas obtenida en un tiempo particular, y sea X el número de las defectuosas y $\hat{p} = X/n$. Debido a que X tiene una distribución binomial, $E(X) = np$ y $V(X) = np(1 - p)$, así

$$E(\hat{p}) = p \qquad V(\hat{p}) = \frac{p(1-p)}{n}$$

También, si $np \geq 10$ y $n(1 - p) \geq 10$, $\hat{p}$ tiene aproximadamente una distribución normal.

En el caso de p conocida (o una gráfica basada en un valor objetivo), los límites de control son

$$\text{Límite inferior de control} = p - 3\sqrt{\frac{p(1-p)}{n}}$$

$$\text{Límite superior de control} = p + 3\sqrt{\frac{p(1-p)}{n}}$$

Si cada muestra está formada por n piezas, el número de las defectuosas de la i-ésima muestra es x_i y $\hat{p}_i = x_i/n$, entonces $\hat{p}_1, \hat{p}_2, \hat{p}_3, \ldots$ se grafican en la gráfica de control.

Generalmente el valor de p debe estimarse a partir de los datos. Suponga que se dispone de k muestras de lo que se piensa que es un proceso bajo control, y que

$$\bar{p} = \frac{\sum_{i=1}^{k} \hat{p}_i}{k}$$

La estimación $\bar{p}$ se usa entonces en lugar de p en los límites de control mencionados líneas antes.

La gráfica p para la fracción de piezas defectuosas tiene su línea de centro a una altura $\bar{p}$ y límites de control

$$\text{LCL} = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

$$\text{UCL} = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

Si el límite inferior de control es negativo, se sustituye con 0.

Ejemplo 16.6 Una muestra de 100 tazas de una figura particular de vajilla se seleccionó en cada uno de 25 días sucesivos, y cada una se examinó en busca de defectos. Los números resultantes de tazas inaceptables y de proporciones muestrales correspondientes son como sigue:

Día (i)	1	2	3	4	5	6	7	8	9	10	11	12	13
x_i	7	4	3	6	4	9	6	7	5	3	7	8	4
$\hat{p}_i$	.07	.04	.03	.06	.04	.09	.06	.07	.05	.03	.07	.08	.04

Día (i)	14	15	16	17	18	19	20	21	22	23	24	25
x_i	6	2	9	7	6	7	11	6	7	4	8	6
$\hat{p}_i$	.06	.02	.09	.07	.06	.07	.11	.06	.07	.04	.08	.06

Suponiendo que el proceso estaba bajo control durante este periodo, establezca límites de control y construya una gráfica p. Se tiene que $\sum \hat{p}_i = 1.52$, $\bar{p} = 1.52/25 = .0608$ y

$$\text{LCL} = .0608 - 3\sqrt{(.0608)(.9392)/100} = .0608 - .0717 = -.0109$$

$$\text{UCL} = .0608 + 3\sqrt{(.0608)(.9392)/100} = .0608 + .0717 = .1325$$

Por tanto, el límite inferior de control está en 0. La gráfica de control ilustrada en la figura 16.5 muestra que todos los puntos están dentro de límites de control. Esto es consistente con la suposición de un proceso bajo control.

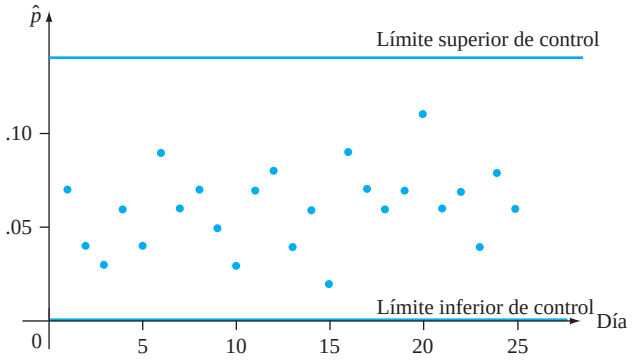


Figura 16.5 Gráfica de control para datos de fracción defectuosa del ejemplo16.6 ■

La gráfica *c* para el número de defectos

A continuación se consideran situaciones en las que la observación de cada punto en el tiempo es el número de defectos en una unidad de algún tipo. La unidad puede consistir en una sola pieza (por ejemplo un automóvil) o un grupo de piezas (por ejemplo, defectos en un conjunto de cuatro llantas). En el segundo caso, se supone que el tamaño del grupo es igual en cada uno de los puntos de tiempo.

La gráfica de control para números de defectos está basada en la distribución de probabilidad Poisson. Recuerde que si Y es una variable aleatoria Poisson con parámetro μ, entonces

$$E(Y) = \mu \qquad V(Y) = \mu \qquad \sigma_Y = \sqrt{\mu}$$

También, Y tiene aproximadamente una distribución normal cuando μ es grande ($\mu \geq 10$ será suficiente para la mayor parte de los casos). Además, si $Y_1, Y_2, \ldots, Y_n$ son variables Poisson independientes con parámetros $\mu_1, \mu_2, \ldots, \mu_n$, se puede demostrar que

$Y_1 + \cdots + Y_n$ tiene una distribución Poisson con parámetro $\mu_1 + \cdots + \mu_n$. En particular, si $\mu_1 = \cdots = \mu_n = \mu$ (la distribución del número de defectos por pieza es igual para cada una de éstas), entonces el parámetro Poisson es $n\mu$.

Denote con μ el parámetro Poisson para el número de defectos en una unidad (es el número esperado de defectos por unidad). En el caso de μ conocida (o una gráfica basada en un valor objetivo),

$$\text{LCL} = \mu - 3\sqrt{\mu} \qquad \text{UCL} = \mu + 3\sqrt{\mu}$$

Con x_i denotando el número total de defectos en la i-ésima unidad ($i = 1, 2, 3, \ldots$), entonces los puntos en las alturas x_1, x_2, x_3, . . . se determinan en la gráfica. Por lo general, el valor de μ debe estimarse a partir de los datos. Como $E(X_i) = \mu$, es natural usar la estimación $\mu = \bar{x}$ (basada en $x_1, x_2, \ldots, x_k$).

La gráfica c para el número de defectos en una unidad tiene línea de centro en $\bar{x}$ y

$$\text{LCL} = \bar{x} - 3\sqrt{\bar{x}}$$

$$\text{UCL} = \bar{x} + 3\sqrt{\bar{x}}$$

Si el límite inferior de control (LCL) es negativo, se sustituye con 0.

Ejemplo 16.7 Una compañía manufactura paneles metálicos que se hornean después de cubrirlos con una mezcla de cerámica en polvo. A veces aparecen fallas en el acabado de estos paneles, y la compañía desea establecer una gráfica de control para determinar el número de fallas. Los números de fallas de cada uno de los 24 paneles muestreados en intervalos regulares son como sigue:

7	10	9	12	13	6	13	7	5	11	8	10
13	9	21	10	6	8	3	12	7	11	14	10

con $\sum x_i = 235$ y $\hat{\mu} = \bar{x} = 235/24 = 9.79$. Los límites de control son

$$\text{LCL} = 9.79 - 3\sqrt{9.79} = .40 \qquad \text{UCL} = 9.79 + 3\sqrt{9.79} = 19.18$$

La gráfica de control está en la figura 16.6 (página 671). El punto correspondiente del 15o. panel está arriba del límite superior de control. Al hacer la investigación, se encontró que la mezcla empleada en ese panel tenía una viscosidad demasiado baja (una causa asignable). La eliminación de esa observación da $\bar{x} = 214/23 = 9.30$ y nuevos límites de control.

$$\text{LCL} = 9.30 - 3\sqrt{9.30} = .15 \qquad \text{UCL} = 9.30 + 3\sqrt{9.30} = 18.45$$

Todas las 23 observaciones restantes están entre estos límites, lo cual indica un proceso bajo control. ■

Gráficas de control basadas en datos transformados

Se presume que el uso de límites de control de 3 sigmas da por resultado P(estadística $<$ límite inferior de control) $\approx$ P(estadística $>$ límite superior de control) $\approx .0013$ cuando el proceso está bajo control. No obstante, cuando p es pequeña, la aproximación normal a la distribución de $\hat{p} = X/n$ con frecuencia no será muy precisa en las colas de extremos. La tabla 16.3 da evidencia de este comportamiento para valores seleccionados de p y n (el valor de p se usa para calcular los límites de control). En muchos casos, la probabilidad de que un solo punto caiga fuera de los límites de control es muy diferente de la probabilidad nominal de .0026.

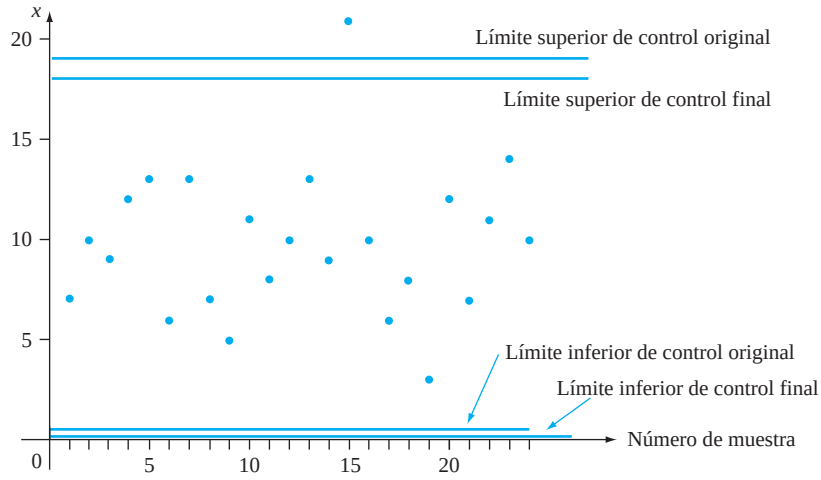


Figura 16.6 Gráfica de control para el número de datos de fallas del ejemplo 16.7

Tabla 16.3 Probabilidades bajo control para una gráfica *p*

p	n	$P(\hat{p} < \text{LCL})$	$P(\hat{p} > \text{UCL})$	P(punto fuera de control)
.10	100	.00003	.00198	.00201
.10	200	.00048	.00299	.00347
.10	400	.00044	.00171	.00215
.05	200	.00004	.00266	.00270
.05	400	.00020	.00207	.00227
.05	600	.00031	.00189	.00220
.02	600	.00007	.00275	.00282
.02	800	.00036	.00374	.00410
.02	1000	.00023	.00243	.00266

Este problema se puede solucionar si se aplica una transformación a los datos. Denote con $h(X)$ una función aplicada para transformar la variable binomial X. Entonces $h(\cdot)$ debe seleccionarse de modo que $h(X)$ tenga aproximadamente una distribución normal y esta aproximación sea precisa en las colas. Una transformación recomendada está basada en la función arcsen (es decir, sen^{-1}):

$$Y = h(X) = \text{sen}^{-1}(\sqrt{X/n})$$

Entonces Y es aproximadamente normal con valor medio $\text{sen}^{-1}(\sqrt{p})$ y varianza $1/(4n)$; nótese que la varianza es independiente de p. Sea $y_i = \text{sen}^{-1}(\sqrt{x_i/n})$. Los puntos en la gráfica de control están localizados a alturas $y_1, y_2, \ldots$. Para n conocida, los límites de control son

$$\text{LCL} = \text{sen}^{-1}(\sqrt{p}) - 3\sqrt{1/(4n)} \quad \text{UCL} = \text{sen}^{-1}(\sqrt{p}) + 3\sqrt{1/(4n)}$$

Cuando p es desconocida, $\text{sen}^{-1}(\sqrt{p})$ se sustituye con $\bar{y}$.

Se aplican comentarios semejantes a la distribución Poisson cuando μ es pequeña. La transformación sugerida es $Y = h(X) = 2\sqrt{X}$, que tiene valor medio $2\sqrt{\mu}$ y varianza 1. Los límites de control resultantes son $2\sqrt{\mu} \pm 3$ cuando μ se conoce y $\bar{y} \pm 3$ cuando no se conoce. El libro *Statistical Methods for Quality Improvement* de la bibliografía del capítulo examina con gran detalle estos problemas.

EJERCICIOS Sección 16.4 (21–28)

21. En cada uno de los 25 días previos, 100 aparatos electrónicos de cierto tipo se seleccionaron al azar y se sometieron a una severa prueba de esfuerzo debido al calor. El número total de piezas que no pasaron la prueba fue de 578.
- **a.** Determine límites de control para una gráfica p de 3 sigmas.
- **b.** El número más alto de piezas que no pasaron la prueba en un día determinado fue 39 y el número más bajo fue 13. ¿Cualquiera de estos números corresponde a un punto fuera de control? Explique.

22. Se seleccionó en cada uno de 30 días consecutivos una muestra de 200 chips de ROM de computadora y el número de chips que no se apegaron a especificaciones en cada día fue como sigue: 10, 18, 24, 17, 37, 19, 7, 25, 11, 24, 29, 15, 16, 21, 18, 17, 15, 22, 12, 20, 17, 18, 12, 24, 30, 16, 11, 20, 14, 28. Construya una gráfica p y examínela en busca de cualquier punto fuera de control.

23. Cuando $n = 150$, ¿cuál es el valor más pequeño de $\bar{p}$ para el que el límite inferior de control de una gráfica p es positivo?

24. Consulte la información del ejercicio 22 y construya una gráfica de control usando la transformación sen^{-1} como se sugirió en el texto.

25. Las observaciones siguientes son números de defectos en 25 especímenes de 1 yarda cuadrada de tela tejida de cierto tipo: 3, 7, 5, 3, 4, 2, 8, 4, 3, 3, 6, 7, 2, 3, 2, 4, 7, 3, 2, 4, 4, 1, 5, 4, 6. Construya una gráfica c para hallar el número de defectos.

26. ¿Para qué valores $\bar{x}$ será negativo el límite inferior de control en una gráfica c?

27. En algunas situaciones, varían los tamaños de especímenes muestreados, y se espera que especímenes más grandes tengan más defectos que los más pequeños. Por ejemplo, los tamaños de muestras de telas inspeccionadas en busca de defectos pueden variar en el tiempo. Alternativamente, el número de piezas inspeccionadas podría cambiar con el tiempo. Sea

$$u_i = \frac{\text{el número de defectos observados en el tiempo } i}{\text{tamaño de entidad inspeccionada en el tiempo } i} = \frac{x_i}{g_i}$$

donde "tamaño" puede referirse a área, longitud, volumen, o simplemente el número de piezas inspeccionadas. Entonces una **gráfica *u*** para $u_1, u_2, \ldots$, tiene línea de centro $\bar{u}$, y los límites de control para las i-ésimas observaciones son $\bar{u} \pm 3\sqrt{\bar{u}/g_i}$.

Se examinaron paneles pintados en secuencia de tiempo y, para cada uno, se determinó el número de defectos en una región de muestreo especificada. El área superficial (en pies cuadrados) de la región examinada varió de panel a panel. Los resultados se dan a continuación. Construya una gráfica u.

Panel	Área examinada	Núm. de defectos
1	.8	3
2	.6	2
3	.8	3
4	.8	2
5	1.0	5
6	1.0	5
7	.8	10
8	1.0	12
9	.6	4
10	.6	2
11	.6	1
12	.8	3
13	.8	5
14	1.0	4
15	1.0	6
16	1.0	12
17	.8	3
18	.6	3
19	.6	5
20	.6	1

28. Construya una gráfica de control para la información del ejercicio 25, mediante el uso de la transformación sugerida en el texto.

16.5 Procedimientos CUSUM

Un defecto de la gráfica X tradicional es su incapacidad para detectar un cambio relativamente pequeño en la media de un proceso. Esto es, en gran medida, consecuencia del hecho de que si un proceso se juzga como fuera de control en un tiempo particular depende sólo de la muestra en ese tiempo, y no de la historia pasada del proceso. Se han diseñado procedimientos y gráficas de control de **suma acumulativa** (CUSUM) para solucionar este defecto.

Hay dos versiones equivalentes de un procedimiento CUSUM para la media de un proceso, una gráfica y otra computacional. Esta última se emplea casi exclusivamente en la práctica, pero la lógica que hay detrás del procedimiento se entiende con más facilidad si se considera primero la forma gráfica.

La mascarilla V

Denote con μ_0 un valor objetivo o meta para la media del proceso, y defina *sumas acumulativas* con

$$S_1 = \bar{x}_1 - \mu_0$$

$$S_2 = (\bar{x}_1 - \mu_0) + (\bar{x}_2 - \mu_0) = \sum_{i=1}^{2}(\bar{x}_i - \mu_0)$$

$$\vdots$$

$$S_l = (\bar{x}_1 - \mu_0) + \cdots + (\bar{x}_l - \mu_0) = \sum_{i=1}^{l}(\bar{x}_i - \mu_0)$$

(en ausencia de un valor objetivo, se usa $\bar{\bar{x}}$ en lugar de μ_0). Estas sumas acumulativas se grafican en el tiempo, es decir, en el tiempo l, se localiza un punto a una altura S_l. En el punto r del tiempo actual, los puntos graficados son $(1, S_1)$, $(2, S_2)$, $(3, S_3)$, . . . , (r, S_r).

En seguida se superpone una "mascarilla" en forma de V sobre la gráfica, como se ve en la figura 16.7. El punto 0, que se encuentra a una distancia d atrás del punto en el que se cruzan los dos brazos de la mascarilla, está posicionado en el actual punto CUSUM (r, S_r). En el tiempo r, el proceso se juzga fuera de control si cualquiera de los puntos graficados se encuentra fuera de la mascarilla V, ya sea arriba del brazo superior o abajo del

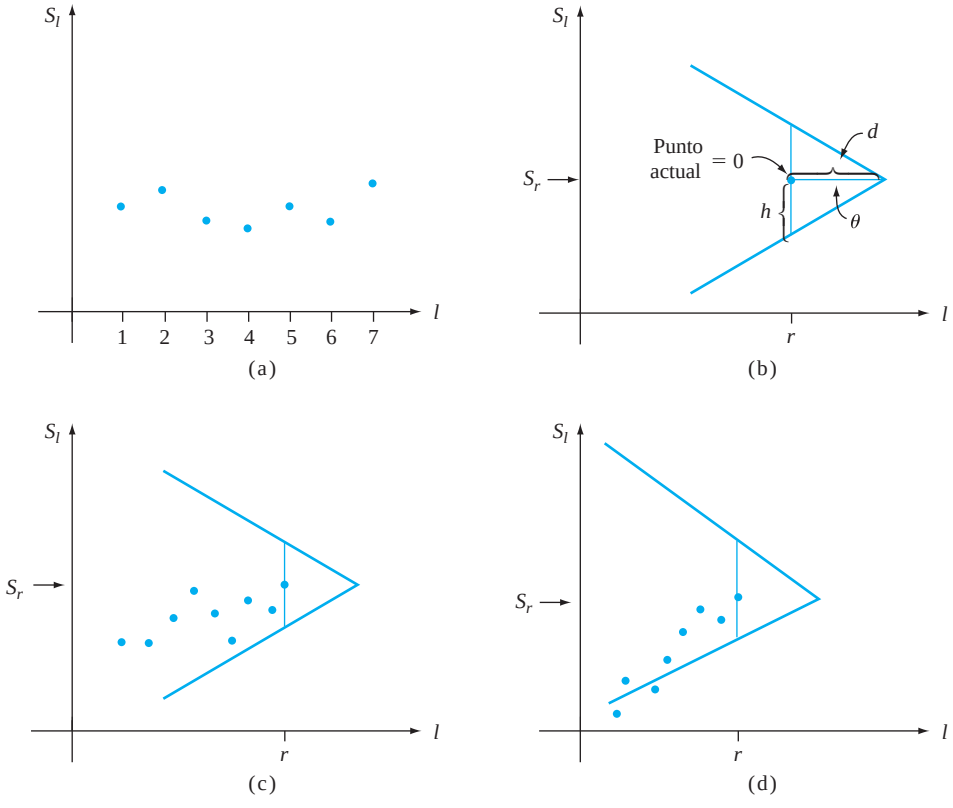


Figura 16.7 Gráficas CUSUM: (a) puntos sucesivos (l, S_l) en una gráfica CUSUM; (b) una mascarilla V con 0 = (r, S_r); (c) un proceso bajo control; (d) un proceso fuera de control

brazo inferior. Cuando el proceso está bajo control, las $\bar{x}_i$ varían alrededor del valor objetivo μ_0, de modo que las S_i sucesivas deben variar alrededor del 0. Suponga, no obstante, que en cierto tiempo, la media del proceso cambia a un valor más grande que el objetivo. De ese punto en adelante, las diferencias $\bar{x}_i - \mu_0$ tenderán a ser positivas, de modo que las S_i sucesivas aumentarán y los puntos graficados se desplazarán hacia arriba. Si ha ocurrido un cambio antes del punto r actual de tiempo, hay una buena probabilidad de que (r, S_r) sea considerablemente más alto que algunos otros puntos de la gráfica, en cuyo caso estos otros puntos estarán debajo del brazo inferior de la mascarilla. Del mismo modo, un cambio a un valor menor que el objetivo dará por resultado subsecuentemente puntos arriba del brazo superior de la mascarilla.

Cualquier mascarilla V particular se determina al especificar la "distancia de avance" d y el "semiángulo" θ, o bien, lo que es equivalente, al especificar d y la longitud h del segmento vertical de recta de 0 al brazo inferior (o al superior) de la mascarilla. Un método para determinar cuál mascarilla usar consiste en especificar el tamaño de un cambio en la media de un proceso que sea de particular interés para un investigador. Entonces los parámetros de la mascarilla se seleccionan para dar valores deseados de α y β, la probabilidad de falsa alarma y la probabilidad de no detectar el cambio especificado, respectivamente. Un método alternativo consiste en seleccionar la mascarilla que dé valores especificados de la magnitud promedio de lote (ARL) para un proceso bajo control y para un proceso en el que la media haya cambiado en una cantidad designada. Después de desarrollar la forma computacional del procedimiento CUSUM, se ilustrará el segundo método de construcción.

Ejemplo 16.8 Una compañía de productos de madera manufactura briquetas de carbón vegetal para barbacoas. Empaca estas briquetas en bolsas de varios tamaños, el más grande de los cuales se supone que contiene 40 libras. La tabla 16.4 muestra los pesos de bolsas de 16 muestras diferentes, cada una de tamaño $n = 4$. Las primeras 10 de éstas se extrajeron de una distribución normal con $\mu = \mu_0 = 40$ y $\sigma = .5$. Comenzando con la muestra 11, la media ha cambiado hacia arriba a $\mu = 40.3$.

Tabla 16.4 Observaciones, $\bar{x}$ y sumas acumulativas para el ejemplo 16.8

Número de muestra	Observaciones				$\bar{x}$	$\sum(\bar{x}_i - 40)$
1	40.77	39.95	40.86	39.21	40.20	.20
2	38.94	39.70	40.37	39.88	39.72	−.08
3	40.43	40.27	40.91	40.05	40.42	.34
4	39.55	40.10	39.39	40.89	39.98	.32
5	41.01	39.07	39.85	40.32	40.06	.38
6	39.06	39.90	39.84	40.22	39.76	.14
7	39.63	39.42	40.04	39.50	39.65	−.21
8	41.05	40.74	40.43	39.40	40.41	.20
9	40.28	40.89	39.61	40.48	40.32	.52
10	39.28	40.49	38.88	40.72	39.84	.36
11	40.57	40.04	40.85	40.51	40.49	.85
12	39.90	40.67	40.51	40.53	40.40	1.25
13	40.70	40.54	40.73	40.45	40.61	1.86
14	39.58	40.90	39.62	39.83	39.98	1.84
15	40.16	40.69	40.37	39.69	40.23	2.07
16	40.46	40.21	40.09	40.58	40.34	2.41

La figura 16.8 muestra una gráfica $\overline{X}$ con límites de control

$$\mu_0 \pm 3\sigma_{\overline{X}} = 40 \pm 3 \cdot (.5/\sqrt{4}) = 40 \pm .75$$

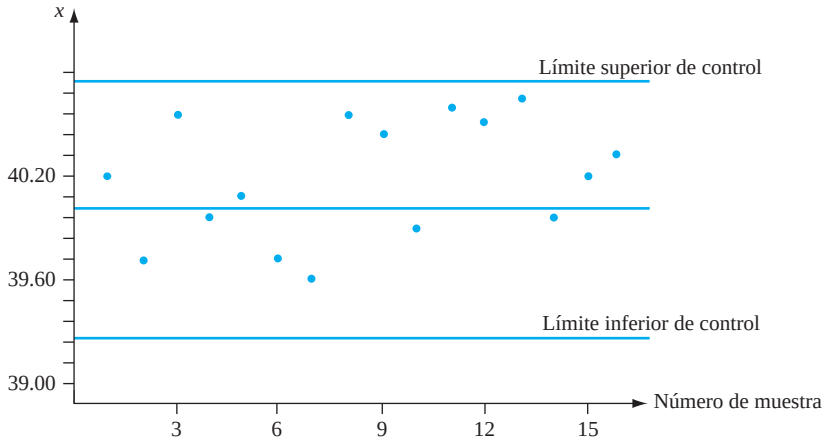


Figura 16.8 Gráfica de control $\overline{X}$ para los datos del ejemplo 16.8

Ningún punto de la gráfica se encuentra fuera de los límites de control. Esta gráfica sugiere un proceso estable para el cual la media ha permanecido en su objetivo.

La figura 16.9 muestra gráficas CUSUM con una mascarilla V particular superpuesta. La gráfica de la figura 16.9(a) es para el tiempo actual $r = 12$. Todos los puntos de esta gráfica están dentro de los brazos de la mascarilla. No obstante, la gráfica para $r = 13$ que se muestra en la figura 16.9(b) da una señal fuera de control. El punto que cae abajo del brazo inferior de la mascarilla sugiere un aumento en el valor de la media del proceso. La mascarilla en $r = 16$ es todavía más enfática en su mensaje de fuera de control. Éste es un marcado contraste respecto a la gráfica $\overline{X}$.

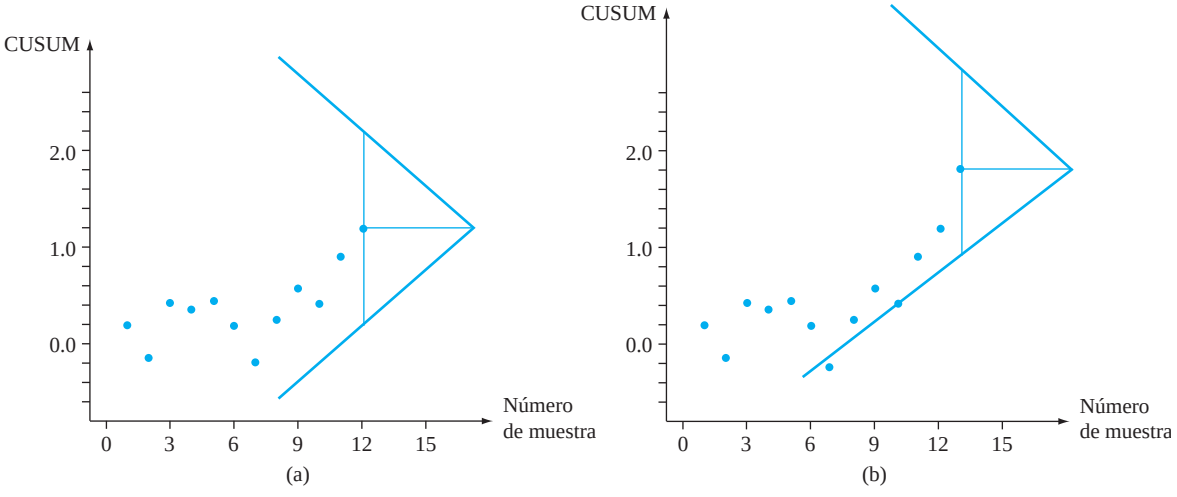


Figura 16.9 Gráficas CUSUM y mascarillas V para datos del ejemplo 16.8: (a) mascarilla V en el tiempo $r = 12$, proceso bajo control; (b) mascarilla V en el tiempo $r = 13$, señal fuera de control ■

Forma computacional

La siguiente versión computacional del procedimiento CUSUM es equivalente a la forma gráfica descrita previamente.

Sea $d_0 = e_0 = 0$, y se calcula $d_1, d_2, d_3, \ldots$ y $e_1, e_2, e_3, \ldots$ en forma repetitiva usando las relaciones

$$d_l = \text{máx}[0, d_{l-1} + (\bar{x}_l - (\mu_0 + k))] \qquad (l = 1, 2, 3, \ldots)$$
$$e_l = \text{máx}[0, e_{l-1} - (\bar{x}_l - (\mu_0 - k))]$$

Aquí el símbolo k denota la pendiente del brazo inferior de la mascarilla V, y su valor se toma cotidianamente como $\Delta/2$ (donde Δ es el tamaño del cambio en μ en el que se concentra la atención).

Si en un tiempo actual r, $d_r > h$ o $e_r > h$, el proceso se juzga como fuera de control. La primera desigualdad sugiere que la media del proceso ha cambiado a un valor mayor que el objetivo, mientras que $e_r > h$ indica un cambio a un valor más pequeño.

Ejemplo 16.9 Reconsidere la información de las briquetas de carbón vegetal que se muestra en la tabla 16.4 del ejemplo 16.8. El valor objetivo es $\mu_0 = 40$, y el tamaño de un cambio a ser rápidamente detectado es $\Delta = .3$. De este modo,

$$k = \frac{\Delta}{2} = .15 \qquad \mu_0 + k = 40.15 \qquad \mu_0 - k = 39.85$$

de modo que

$$d_l = \text{máx}[0, d_{l-1} + (\bar{x}_l - 40.15)]$$
$$e_l = \text{máx}[0, e_{l-1} - (\bar{x}_l - 39.85)]$$

El cálculo de las primeras d_l continúa como sigue:

$$\begin{aligned} d_0 &= 0 \\ d_1 &= \text{máx}[0, d_0 + (\bar{x}_1 - 40.15)] \\ &= \text{máx}[0, 0 + (40.20 - 40.15)] \\ &= .05 \\ d_2 &= \text{máx}[0, d_1 + (\bar{x}_2 - 40.15)] \\ &= \text{máx}[0, .05 + (39.72 - 40.15)] \\ &= 0 \\ d_3 &= \text{máx}[0, d_2 + (\bar{x}_3 - 40.15)] \\ &= \text{máx}[0, 0 + (40.42 - 40.15)] \\ &= .27 \end{aligned}$$

Los demás cálculos se resumen en la tabla 16.5.

El valor $h = .95$ da un procedimiento CUSUM con propiedades deseables, es decir, las alarmas falsas (señales fuera de control incorrectas) se presentan rara vez, pero un cambio de $\Delta = .3$ por lo general será detectado rápidamente. Con este valor de h, la primera señal fuera de control proviene después que se disponga de la 13a. muestra. Como $d_{13} =$

Tabla 16.5 Cálculos de CUSUM para el ejemplo 16.9

Número de muestra	$\bar{x}_l$	$\bar{x}_l - 40.15$	d_l	$\bar{x}_l - 39.85$	e_l
1	40.20	.05	.05	.35	0
2	39.72	−.43	0	−.13	.13
3	40.42	.27	.27	.57	0
4	39.98	−.17	.10	.13	0
5	40.06	−.09	.01	.21	0
6	39.76	−.39	0	−.09	.09
7	39.65	−.50	0	−.20	.29
8	40.41	.26	.26	.56	0
9	40.32	.17	.43	.47	0
10	39.84	−.31	.12	−.01	.01
11	40.49	.34	.46	.64	0
12	40.40	.25	.71	.55	0
13	40.61	.46	1.17	.76	0
14	39.98	−.17	1.00	.13	0
15	40.23	.08	1.08	.38	0
16	40.34	.19	1.27	.49	0

$1.17 > .95$, parece que la media ha cambiado a un valor mayor que el objetivo. Éste es el mismo mensaje que el proporcionado por la mascarilla V de la figura 16.9(b). ■

Para demostrar la equivalencia, de nueva cuenta denote con r el punto actual de tiempo, de modo que $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_r$ están disponibles. La figura 16.10 muestra una mascarilla V con el punto marcado como 0 en (r, S_r). La pendiente del brazo inferior, que se denota por k, es h/d. De este modo, los puntos sobre el brazo inferior arriba de r, $r - 1$, $r - 2$, . . . están a alturas $S_r - h$, $S_r - h - k$, $S_r - h - 2k$, y así sucesivamente.

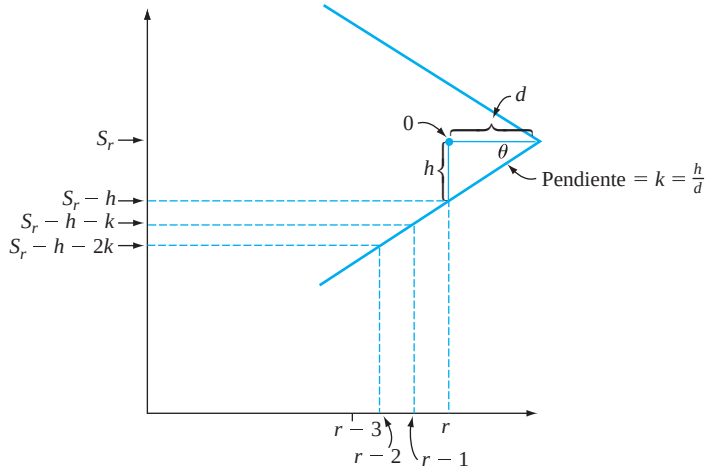


Figura 16.10 Mascarilla V con pendiente de brazo inferior = k

El proceso está bajo control si todos los puntos caen o están entre los brazos de la mascarilla. Para describir esta condición algebraicamente, sea

$$T_l = \sum_{i=1}^{l} [\bar{x}_i - (\mu_0 + k)] \qquad l = 1, 2, 3, \ldots, r$$

Las condiciones bajo las que todos los puntos están sobre o arriba del brazo inferior son

$$\begin{aligned} S_r - h &\le S_r && \text{(trivialmente satisfecha), es decir, } S_r \le S_r + h \\ S_r - h - k &\le S_{r-1} && \text{es decir, } S_r \le S_{r-1} + h + k \\ S_r - h - 2k &\le S_{r-2} && \text{es decir, } S_r \le S_{r-2} + h + 2k \\ &\vdots && \vdots \end{aligned}$$

Ahora se resta rk de ambos lados de cada desigualdad para obtener

$$\begin{aligned} S_r - rk &\le S_r - rk + h && \text{es decir, } T_r \le T_r + h \\ S_r - rk &\le S_{r-1} - (r-1)k + h && \text{es decir, } T_r \le T_{r-1} + h \\ S_r - rk &\le S_{r-2} - (r-2)k + h && \text{es decir, } T_r \le T_{r-2} + h \\ &\vdots && \vdots \end{aligned}$$

Entonces todos los puntos graficados se encuentran sobre o arriba del brazo inferior si y sólo si $T_r - T_r \le h$, $T_r - T_{r-1} \le h$, $T_r - T_{r-2} \le h$ y así sucesivamente. Esto es equivalente a

$$T_r - \text{mín}(T_1, T_2, \ldots, T_r) \le h$$

De modo semejante, si se hace

$$V_r = \sum_{i=1}^{r} [\bar{x}_i - (\mu_0 - k)] = S_r + rk$$

se puede demostrar que todos los puntos se encuentran sobre o abajo del brazo superior si y sólo si

$$\text{máx}(V_1, \ldots, V_r) - V_r \le h$$

Si ahora se hace

$$\begin{aligned} d_r &= T_r - \text{mín}(T_1, \ldots, T_r) \\ e_r &= \text{máx}(V_1, \ldots, V_r) - V_r \end{aligned}$$

se ve fácilmente que $d_1, d_2, \ldots$ y $e_1, e_2, \ldots$ se pueden calcular de manera repetitiva como ya se ilustró aquí líneas antes. Por ejemplo, la expresión para d_r de acuerdo a la consideración de dos casos:

1. $\text{mín}(T_1, \ldots, T_r) = T_r$, de donde $d_r = 0$

2. $\text{mín}(T_1, \ldots, T_r) = \text{mín}(T_1, \ldots, T_{r-1})$, de modo que

$$\begin{aligned} d_r &= T_r - \text{mín}(T_1, \ldots, T_{r-1}) \\ &= \bar{x}_r - (\mu_0 + k) + T_{r-1} - \text{mín}(T_1, \ldots, T_{r-1}) \\ &= \bar{x}_r - (\mu_0 + k) + d_{r-1} \end{aligned}$$

Como d_r no puede ser negativa, es la mayor de estas dos cantidades.

Diseño de un procedimiento CUSUM

Denote con Δ el tamaño de un cambio en μ que ha de ser detectado rápidamente usando un procedimiento CUSUM.* Es práctica común hacer $k = \Delta/2$. Ahora suponga que un experto en control de calidad especifica valores deseados de dos grandes magnitudes promedio de lote:

* Esto contrasta con la anotación previa, donde Δ representaba el número de desviaciones estándar por las que cambiaba μ.

1. ARL (magnitud promedio de lote) cuando el proceso está bajo control ($\mu = \mu_0$)
2. ARL (magnitud promedio de lote) cuando el proceso está fuera de control porque la media ha cambiado en Δ ($\mu = \mu_0 + \Delta$ o bien $\mu = \mu_0 - \Delta$)

Una gráfica desarrollada por Kenneth Kemp ("The Use of Cumulative Sums for Sampling Inspection Schemes", *Applied Statistics*, 1962: 23), llamada *nomograma*, se puede usar entonces para determinar valores de h y n que alcancen las magnitudes promedio de lote especificadas.[†] El método para usar la gráfica se describe en el recuadro siguiente.
El valor de σ debe ser conocido o se usa una estimación en su lugar.

Uso del nomograma de Kemp

1. Localice las ARL en las escalas bajo control y fuera de control. Enlace estos dos puntos con una recta.
2. Observe en dónde la recta cruza la escala k', y despeje n usando la ecuación

$$k' = \frac{\Delta/2}{\sigma/\sqrt{n}}$$

   A continuación redondee n hacia arriba al entero más próximo.
3. Enlace el punto en la escala k' con el punto de la escala ARL bajo control usando una segunda recta, y observe en dónde esta recta cruza la escala h'. Entonces $h = (\sigma/\sqrt{n}) \cdot h'$.

El valor $h = .95$ se usó en el ejemplo 16.9. En esa situación, se deduce que la ARL bajo control es 500 y la ARL fuera de control (para $\Delta = .3$) es 7.

Ejemplo 16.10 El valor objetivo para el diámetro del núcleo interior de una bomba hidráulica es 2.250 pulgadas. Si la desviación estándar del diámetro del núcleo es $\sigma = .004$, ¿cuál procedimiento CUSUM dará una ARL bajo control de 500 y una ARL de 5 cuando el diámetro medio del núcleo cambie en una cantidad de .003 pulgada?

Si se enlaza el punto 500 de una escala ARL bajo control con el punto 5 de la escala ARL fuera de control, y se prolonga la recta a la escala k' de la extrema izquierda de la figura 16.11, resulta $k' = .74$. Por tanto,

$$k' = .74 = \frac{\Delta/2}{\sigma/\sqrt{n}} = \frac{.0015}{.004/\sqrt{n}} = .375\sqrt{n}$$

de modo que

$$\sqrt{n} = \frac{.74}{.375} = 1.973 \qquad n = (1.973)^2 = 3.894$$

Por tanto, el procedimiento CUSUM debe estar basado en el tamaño muestral $n = 4$. Ahora, si se enlaza .74 en la escala k' con 500 en la escala ARL bajo control resulta $h' = 3.2$, de donde

$$h = (\sigma/\sqrt{n}) \cdot (3.2) = (.004/\sqrt{4})(3.2) = .0064$$

Resulta una señal fuera de control tan pronto como $d_r > .0064$ o $e_r > .0064$. ■

Se han examinado procedimientos CUSUM para controlar la ubicación de un proceso. Existen también procedimientos CUSUM para controlar variaciones en procesos y para datos de atributos. Deben consultarse las referencias del capítulo para ver información sobre estos procedimientos.

EJERCICIOS Sección 16.5 (29–32)

29. Se supone que los recipientes de cierto tratamiento para fosas sépticas contienen 16 onzas de líquido. Se selecciona una muestra de cinco recipientes de la línea de producción, una vez por hora, y se determina el contenido promedio de la muestra. Considere los siguientes resultados: 15.992, 16.051, 16.066, 15.912, 16.030, 16.060, 15.982, 15.899, 16.038, 16.074, 16.029, 15.935, 16.032, 15.960, 16.055. Usando $\Delta = .10$ y $h = .20$, utilice la forma computacional del procedimiento CUSUM para investigar el comportamiento de este proceso.

30. El valor objetivo para el diámetro de cierto tipo de eje motor es de .75 pulgada. El tamaño del cambio en el diámetro promedio considerado importante a detectar es .002 pulgada. Haga una muestra de diámetros promedio para grupos sucesivos de $n = 4$ ejes como sigue: .7507, .7504, .7492, .7501, .7503, .7510, .7490, .7497, .7488, .7504, .7516, .7472, .7489, .7483, .7471, .7498, .7460, .7482, .7470, .7493, .7462, .7481. Utilice la forma computacional del procedimiento CUSUM con $h = .003$ para ver si la media del proceso permaneció en su objetivo durante todo el tiempo de observación.

31. La desviación estándar de cierta dimensión en una pieza de avión es de .005 cm. ¿Qué procedimiento CUSUM dará una ARL bajo control de 600 y una ARL fuera de control de 4 cuando el valor medio de la dimensión cambia en .004 cm?

32. Cuando la ARL fuera de control (magnitud promedio de lote) corresponde a un cambio de 1 desviación estándar en la media de un proceso, ¿cuáles son las características del procedimiento CUSUM que tiene unas ARL de 250 y 4.8 para las condiciones bajo control y fuera de control, respectivamente?

16.6 Muestreo de aceptación

Es frecuente que las piezas que provienen de un proceso de producción se envíen en grupos a otra compañía o establecimiento comercial. Un grupo podría estar formado por todas las unidades de un lote o turno de producción particular, en un contenedor de embarque de algún tipo, enviado en respuesta a un pedido en particular y así sucesivamente. El grupo

de piezas suele recibir el nombre de *lote*, el remitente se conoce como *productor* y el destinatario del lote es el *consumidor*. Aquí el interés se centrará en situaciones en las que cada pieza es defectuosa o no defectuosa, con p denotando la proporción de unidades defectuosas del lote. Naturalmente que el consumidor desearía aceptar el lote sólo si el valor de p es adecuadamente pequeño. El muestreo de aceptación es la parte de la estadística aplicada que se refiere a métodos para determinar si el consumidor debería aceptar o rechazar un lote.

Hasta hace poco tiempo, los especialistas consideraban que los procedimientos de gráficas de control y técnicas de muestreo de aceptación eran partes igualmente importantes de la metodología de control de calidad. Éste ya no es el caso. La razón es que el uso de gráficas de control y otras estrategias creadas en tiempos recientes ofrece la oportunidad de diseñar calidad en un producto, mientras que el muestreo de aceptación resuelve lo que ya se ha producido y no estipula ningún control directo sobre la calidad del proceso. Esto llevó al desaparecido experto en control de calidad, el norteamericano W. E. Deming, especialista importante, a persuadir a los japoneses de hacer el mayor uso posible de la metodología de control de calidad, a razonar firmemente contra el uso del muestreo de aceptación en muchas situaciones. En un estilo semejante, el reciente libro de Ryan (vea la bibliografía del capítulo) dedica varios capítulos a gráficas de control y menciona el muestreo de aceptación sólo de paso. Como reflejo de esto, se presenta aquí una breve introducción a conceptos básicos.

Planes de un solo muestreo

El tipo más claro de plan de muestreo de aceptación consiste en seleccionar una sola muestra aleatoria de tamaño n, y a continuación rechazar el lote si el número de defectos en la muestra excede de un valor c crítico especificado. Denote con la variable aleatoria X el número de piezas defectuosas del lote, y con A el evento de que el lote se acepte. Entonces $P(A) = P(X \le c)$ es una función de p; cuanto mayor sea el valor de p, menor será la probabilidad de aceptar el lote.

Si el tamaño muestral n es grande con relación a N, $P(A)$ se calcula usando la distribución hipergeométrica (el número de piezas defectuosas en el lote es Np):

$$P(X \le c) = \sum_{x=0}^{c} h(x; n, Np, N) = \sum_{x=0}^{c} \frac{\binom{Np}{x} \cdot \binom{N(1-p)}{n-x}}{\binom{N}{n}}$$

Cuando n es pequeña en relación con N (la regla empírica sugerida previamente fue $n \le .05N$, pero algunos autores emplean la regla menos conservadora $n \le .10/N$), se puede usar la distribución binomial:

$$P(X \le c) = \sum_{x=0}^{c} b(x; n, p) = \sum_{x=0}^{c} \binom{n}{x} p^x (1-p)^{n-x}$$

Por último, si $P(A)$ es grande sólo cuando p es pequeña (esto depende del valor de c), se justifica la aproximación Poisson a la distribución binomial:

$$P(X \le c) \approx \sum_{x=0}^{c} p(x; np) = \sum_{x=0}^{c} \frac{e^{-np}(np)^x}{x!}$$

El comportamiento de un plan de muestreo se puede resumir muy bien si se grafica $P(A)$ como una función de p. Esa gráfica se denomina **curva característica de operación (OC)** para el plan.

Ejemplo 16.11 Considere el plan de muestreo con $c = 2$ y $n = 50$. Si el tamaño N del lote excede de 1000, se puede usar la distribución binomial. Esto da

$$P(A) = P(X \le 2) = (1-p)^{50} + 50p(1-p)^{49} + 1255p^2(1-p)^{48}$$

La tabla siguiente muestra $P(A)$ para valores seleccionados de p, y la correspondiente curva característica de operación se muestra en la figura 16.12.

p	.01	.02	.03	.04	.05	.06	.07	.08	.09	.10	.12	.15
$P(A)$	.986	.922	.811	.677	.541	.416	.311	.226	.161	.112	.051	.014

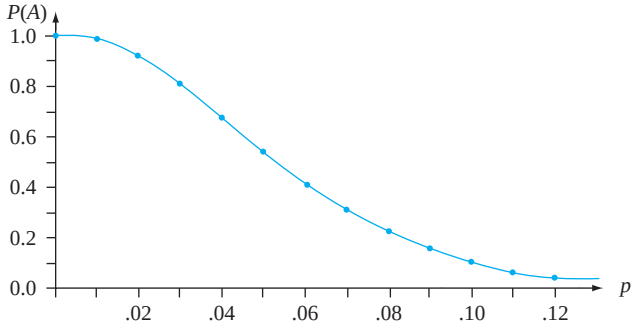


Figura 16.12 Curva característica de operación para plan de muestreo con $c = 2, n = 50$ ■

La curva característica de operación para el plan del ejemplo 16.11 tiene $P(A)$ cerca de 1 para p muy cercana a 0. No obstante, en numerosas aplicaciones un porcentaje de piezas defectuosas de 8% [para el que $P(A) = .226$] o incluso sólo 5% [$P(A) = .541$] sería considerado excesivo, en cuyo caso las probabilidades de aceptación son demasiado altas. Si se aumenta el valor crítico c y se mantiene n fija da por resultado un plan para el que $P(A)$ aumenta en cada p (excepto 0 y 1), de modo que la nueva curva característica de operación se encuentra arriba de la anterior. Esto es deseable para p cercana a 0 pero no para valores más grandes de p. Al mantener c constante mientras se aumenta n se obtiene una curva característica de operación más baja, que está bien para p más grande pero no para p cercana a 0. Se desea una curva característica de operación que sea más alta para una p muy pequeña y más baja para una p más grande. Esto se logra al aumentar n y ajustar c.

Diseño de un plan de una muestra

Un plan de muestreo eficaz es aquel con las siguientes características:

1. Tiene una alta probabilidad especificada de aceptar lotes que el productor considera que son de buena calidad.
2. Tiene una baja probabilidad especificada de aceptar lotes que el consumidor considera que son de mala calidad.

Un plan de este tipo se puede llevar a cabo si se procede como sigue. Se designan dos valores diferentes de p, uno para el que $P(A)$ es un valor especificado cercano a 1 y el otro para el que $P(A)$ es un valor especificado cercano a 0. Estos dos valores de p, p_1 y p_2 por ejemplo, reciben el nombre de **nivel de calidad aceptable** (AQL) y el **porcentaje defectuoso de tolerancia del lote** (LTPD). Esto es, se necesita un plan para el que

1. $P(A) = 1 - \alpha$ cuando $p = p_1 =$ AQL (α pequeña)
2. $P(A) = \beta$ cuando $p = p_2 =$ LTPD (β pequeña)

Esto es análogo a buscar un procedimiento de prueba con probabilidad α de error tipo I especificado, y probabilidad β de error tipo II especificado cuando se prueben hipótesis. Por ejemplo, se podría tener

$$\text{AQL} = .01 \quad \alpha = .05 \quad (P(A) = .95)$$
$$\text{LTPD} = .045 \quad \beta = .10 \quad (P(A) = .10)$$

Debido a que X es discreta, por lo general hay que contentarse con valores de n y c que aproximadamente satisfagan estas condiciones.

La tabla 16.6 da información de la cual n y c se puedan determinar en el caso $\alpha = .05$, $\beta = .10$.

Tabla 16.6 Factores para determinar n y c para un plan de una muestra con $\alpha = .05$, $\beta = .10$.

c	np_1	np_2	p_2/p_1	c	np_1	np_2	p_2/p_1
0	.051	2.30	45.10	8	4.695	12.99	2.77
1	.355	3.89	10.96	9	5.425	14.21	2.62
2	.818	5.32	6.50	10	6.169	15.41	2.50
3	1.366	6.68	4.89	11	6.924	16.60	2.40
4	1.970	7.99	4.06	12	7.690	17.78	2.31
5	2.613	9.28	3.55	13	8.464	18.86	2.24
6	3.285	10.53	3.21	14	9.246	20.13	2.18
7	3.981	11.77	2.96	15	10.040	21.29	2.12

Ejemplo 16.12 Determine un plan para el que AQL $= p_1 = .01$ y LTPD $= p_2 = .045$. La razón entre p_2 y p_1 es

$$\frac{\text{LTPD}}{\text{AQL}} = \frac{p_2}{p_1} = \frac{.045}{.01} = 4.50$$

Este valor se encuentra entre la razón 4.89 dada en la tabla 16.6, para la cual $c = 3$, y 4.06, para la cual $c = 4$. Una vez que se seleccione uno de estos valores de c, n se puede determinar ya sea al dividir el valor np_1 de la tabla 16.6 entre p_1, o por medio de np_2/p_2. De esta manera cuatro planes diferentes (dos valores de c, y uno por cada uno de los dos valores de n) dan aproximadamente el valor especificado de α y β. Considere, por ejemplo, usar $c = 3$ y

$$n = \frac{np_1}{p_1} = \frac{1.366}{.01} = 136.6 \approx 137$$

Entonces

$$\alpha = 1 - P(X \le 3 \text{ cuando } p = p_1) = .050$$

(la aproximación Poisson con $\mu = 1.37$ también da .050) y

$$\beta = P(X \le 3 \text{ cuando } p = p_2) = .131$$

El plan con $c = 4$ y n determinada a partir de $np_2 = 7.99$ tiene $n = 178$, $\alpha = .034$, y $\beta = .094$. El mayor tamaño muestral da por resultado un plan con α y β menores que los correspondientes valores especificados. ■

El libro de Douglas Montgomery citado en la bibliografía del capítulo contiene una gráfica de la cual c y n se pueden determinar para *cualesquiera* α y β especificadas.

Puede ocurrir que el número de piezas defectuosas de la muestra llegue a $c + 1$ antes que se hayan examinado todas las piezas. Por ejemplo, en el caso $c = 3$ y $n = 137$, puede ser que la pieza número 125 examinada sea la cuarta pieza defectuosa, de modo que no hay necesidad de examinar las 12 piezas restantes. No obstante, por lo general se recomienda examinar todas las piezas incluso cuando esto ocurra para obtener un historial de calidad lote por lote, así como estimaciones de p en el tiempo.

Planes de muestreo doble

En un plan de muestreo doble, se determina el número de piezas defectuosas x_1 de una muestra inicial de tamaño n_1. Hay entonces tres cursos de acción posibles: aceptar de inmediato el lote, rechazar de inmediato el lote, o tomar una segunda muestra de n_2 piezas y rechazar o aceptar el lote dependiendo del número total $x_1 + x_2$ de piezas defectuosas en las dos muestras. Además de los dos tamaños muestrales, un plan específico está caracterizado por otros tres números: c_1, r_1 y c_2 como sigue:

1. Rechazar el lote si $x_1 \geq r_1$.
2. Aceptar el lote si $x_1 \leq c_1$.
3. Si $c_1 < x_1 < r_1$, tomar una segunda muestra; entonces aceptar el lote si $x_1 + x_2 \leq c_2$ y rechazarlo de otro modo.

Ejemplo 16.13 Considere el plan de muestreo doble con $n_1 = 80$, $n_2 = 80$, $c_1 = 2$, $r_1 = 5$, y $c_2 = 6$. De este modo, el lote será aceptado si (1)$x_1 = 0$, 1, o 2; (2)$x_1 = 3$ y $x_2 = 0$, 1, 2, o 3; o (3)$x_1 = 4$ y $x_2 = 0$, 1, o 2.

Suponiendo que el tamaño del lote es grande lo suficiente para que aplique la aproximación binomial, la probabilidad $P(A)$ de aceptar el lote es

$$\begin{aligned} P(A) &= P(X_1 = 0, 1, \text{ o } 2) + P(X_1 = 3,\ X_2 = 0, 1, 2, \text{ o } 3) \\ &\quad + P(X_1 = 4, X_2 = 0, 1, \text{ o } 2) \\ &= \sum_{x_1=0}^{2} b(x_1; 80, p) + b(3; 80, p) \sum_{x_2=0}^{3} b(x_2; 80, p) \\ &\quad + b(4; 80, p) \sum_{x_2=0}^{2} b(x_2; 80, p) \end{aligned}$$

De nuevo la gráfica de $P(A)$ en función de p es la curva característica de operación del plan. La curva característica de operación para este plan aparece en la figura 16.13.

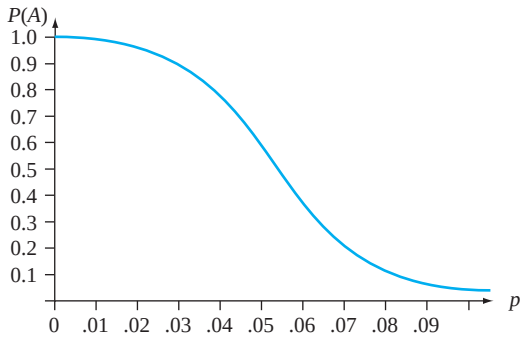


Figura 16.13 Curva característica de operación para el plan de muestreo doble del ejemplo 16.13

Un método común para diseñar un plan de muestreo doble consiste en proceder como se sugirió líneas antes para planes de una sola muestra. Especifique valores p_1 y p_2 junto con sus correspondientes probabilidades de aceptación $1 - \alpha$ y β. A continuación encuentre un plan que satisfaga estas condiciones. El libro de Montgomery contiene tablas semejantes a la tabla 16.6 para este propósito en los casos $n_2 = n_1$ y $n_2 = 2n_1$ con $1 - \alpha = .95$, $\beta = .10$. En otras fuentes se pueden hallar tabulaciones de planes mucho más extensas.

De modo análogo a la práctica común con planes de una sola muestra, se recomienda que todas las piezas de la primera muestra sean examinadas incluso cuando la $(r_1 + 1)$-ésima pieza defectuosa se descubra antes de inspeccionar la n_1-ésima muestra. No obstante, se acostumbra terminar la inspección de la segunda muestra si el número de piezas defectuosas es suficiente para justificar el rechazo antes que se hayan examinado todas las piezas. Esto se conoce como *acortamiento* en la segunda muestra. Con el acortamiento, se puede demostrar que el número esperado de piezas inspeccionadas, en un plan de muestreo doble, es menor que el número de piezas examinadas en un plan de muestreo de una sola pieza, cuando las curvas características de operación de los dos planes están cerca de ser idénticas. Ésta es la principal virtud de planes de muestreo doble. Para saber más de estos temas, así como de un examen de planes de muestreo secuenciales y múltiples (que consisten en seleccionar artículos para inspección uno por uno en lugar de en grupos), debe consultarse un libro que hable de control de calidad.

Rectificación de inspección y otros criterios de diseño

En algunas situaciones, una inspección de muestreo se lleva a cabo mediante *rectificación*. Para planes de una sola muestra, esto significa que cada pieza defectuosa de la muestra se sustituye con una satisfactoria, y si el número de piezas defectuosas de la muestra excede del corte c de aceptación, *todas* las piezas del lote son examinadas y con piezas buenas se sustituye cualquiera de las defectuosas. Denótese con N el tamaño del lote. Una característica importante de un plan de muestreo con rectificación de inspección es la **calidad de salida promedio,** denotada por AOQ. Ésta es la proporción a largo plazo de piezas defectuosas entre las enviadas después de que se utiliza el plan de muestreo. Ahora las piezas defectuosas se presentarán sólo entre las $N - n$ piezas no inspeccionadas de un lote juzgado aceptable con base en una muestra. Supóngase por ejemplo, que $P(A) = P(X \leq c) = .985$ cuando $p = .01$. Entonces, a largo plazo, 98.5% de las $N - n$ piezas que no estén en la muestra no se inspeccionarán, de las que se espera que 1% sean defectuosas. Esto implica que el número esperado de piezas defectuosas de un lote seleccionado al azar sea $(N - n) \cdot P(A) \cdot p = .00985(N - n)$. Si se divide esto entre el número de piezas de un lote resulta la calidad de salida promedio:

$$\begin{aligned} \text{AOQ} &= \frac{(N - n) \cdot P(A) \cdot p}{N} \\ &\approx P(A) \cdot p \qquad \text{si } N \gg n \end{aligned}$$

Debido a que AOQ $= 0$ cuando $p = 0$ o cuando $p = 1$ [$P(A) = 0$ en el último caso], se deduce que hay un valor de p entre 0 y 1 para el cual AOQ es máxima. El valor máximo de AOQ se denomina **límite de calidad de salida promedio, AOQL.** Por ejemplo, para el plan con $n = 137$ y $c = 3$ examinado antes, AOQL $= .0142$, el valor de AOQ en $p \approx .02$.

Las selecciones apropiadas de n y c darán un plan de muestreo para el que AOQL es un número especificado pequeño. No obstante, este plan no es único y se pueden imponer otras condiciones. Es frecuente que esta segunda condición comprenda el **promedio** (es decir, esperado) del **número total inspeccionado,** denotado por **ATI.** El número de piezas inspeccionadas de un lote seleccionado al azar es una variable aleatoria que toma el valor n con probabilidad $P(A)$ y N con probabilidad $1 - P(A)$. Por tanto, el número esperado de piezas inspeccionadas de un lote seleccionado al azar es

$$\text{ATI} = n \cdot P(A) + N \cdot (1 - P(A))$$

Es práctica común seleccionar un plan de muestreo que tenga un límite de calidad de salida promedio (AOQL) y, además, el mínimo del número total inspeccionado (ATI) a un nivel de calidad particular p.

Planes de muestro estándar

Puede parecer como si la determinación de un plan de muestreo que simultáneamente satisfaga varios criterios fuera muy difícil. Por fortuna, ya otros investigadores han puesto las bases en forma de extensas tabulaciones de esos planes. MIL STD 105D, creado por las fuerzas militares después de la Segunda Guerra Mundial, es el conjunto de planes de más amplio uso. Una versión civil, ANSI/ASQC Z1.4, es muy semejante a la versión militar. Un tercer conjunto de planes que goza de preferencia fue creado por los Laboratorios Bell antes de la Segunda Guerra Mundial por dos expertos en estadística, Dodge y Romig. El libro de Montgomery (vea la bibliografía del capítulo) contiene una amena introducción al uso de estos planes.

EJERCICIOS Sección 16.6 (33–40)

33. Considere el plan de una sola muestra con $c = 2$ y $n = 50$, como se vio en el ejemplo 16.11, pero ahora suponga que el tamaño del lote es $N = 500$. Calcule $P(A)$, la probabilidad de aceptar el lote, para $p = .01, .02, \ldots, .10$ usando la distribución hipergeométrica. ¿La aproximación binomial da resultados satisfactorios en este caso?

34. Ha de seleccionarse una muestra de 50 piezas de un lote formado por 5000 piezas. El lote será aceptado si la muestra contiene una pieza defectuosa, a lo sumo. Calcule la probabilidad de aceptación del lote para $p = .01, .02, \ldots, .10$, y trace la curva característica de operación.

35. Consulte el ejercicio 34 y considere el plan con $n = 100$ y $c = 2$. Calcule $P(A)$ para $p = .01, .02, \ldots, .05$, y trace las dos curvas características de operación en el mismo conjunto de ejes. ¿Cuál de los dos planes es preferible (dejando de lado el costo del muestreo) y por qué?

36. Elabore un plan de una sola muestra para el que el nivel de calidad aceptable = .02 y el porcentaje defectuoso de tolerancia del lote = .07 en el caso $\alpha = .05$, $\beta = .10$. Una vez que los valores de n y c se hayan determinado, calcule los valores alcanzados de α y β para el plan.

37. Considere el plan de muestreo doble para el cual ambos tamaños muestrales son de 50. El lote es aceptado después de la primera muestra si el número de piezas defectuosas es 1 a lo sumo, rechazado si el número de piezas defectuosas es al menos 4, y rechazado después de la segunda muestra si el número total de piezas defectuosas es 6 o más. Calcule la probabilidad de aceptar el lote cuando $p = .01, .05$ y $.10$.

38. Algunas fuentes están a favor de un tipo un poco más restrictivo de plan de muestreo doble en el que $r_1 = c_2 + 1$; es decir, el lote es rechazado si en cualquiera de las etapas el número (total) de piezas defectuosas es al menos r_1 (vea el libro de Montgomery). Considere este tipo de plan de muestreo con $n_1 = 50$, $n_2 = 100$, $c_1 = 1$, y $r_1 = 4$. Calcule la probabilidad de aceptación del lote cuando $p = .02, .05$, y $.10$.

39. Consulte el ejemplo 16.11, en el que se utilizó un plan de una sola muestra con $n = 50$ y $c = 2$.
- **a.** Calcule la calidad de salida promedio para $p = .01, .02, \ldots, .10$. ¿Qué sugiere esto acerca del valor de p para el que la calidad de salida promedio es máxima y el correspondiente límite de calidad de salida promedio?
- **b.** Determine el valor de p para el que la calidad de salida promedio sea máxima y el valor correspondiente de AOQL. [*Sugerencia*: use cálculo.]
- **c.** Para $N = 2000$, calcule el promedio del número total inspeccionado para los valores de p dados en el inciso (a).

40. Considere el plan de una sola muestra que utiliza $n = 50$ y $c = 1$ cuando $N = 2000$. Determine los valores de la calidad de salida promedio y el promedio del número total inspeccionado para valores seleccionados de p, y grafique cada uno de éstos contra p. También determine el valor del límite de calidad de salida promedio.

EJERCICIOS SUPLEMENTARIOS (41–46)

41. Observaciones realizadas en la resistencia al corte de 26 subgrupos de soldadura eléctrica por puntos de prueba, cada uno formado por seis soldaduras, producen $\sum \bar{x}_i = 10{,}980$, $\sum s_i = 402$ y $\sum r_i = 1074$. Calcule límites de control para cualquier gráfica de control relevante.

42. Se determina el número de arañazos en la superficie de cada una de 24 placas metálicas rectangulares, dando la siguiente información: 8, 1, 7, 5, 2, 0, 2, 3, 4, 3, 1, 2, 5, 7, 3, 4, 6, 5, 2, 4, 0, 10, 2, 6. Construya una gráfica de control apropiada y comente.

43. Los siguientes números son observaciones sobre resistencia a la tensión de especímenes de tela sintética seleccionados de un proceso de producción a intervalos igualmente espaciados. Construya gráficas de control apropiadas, y comente (suponga que una causa asignable es identificable para cualesquiera observaciones fuera de control).

1.	51.3	51.7	49.5	**12.**	49.6	48.4	50.0
2.	51.0	50.0	49.3	**13.**	49.8	51.2	49.7
3.	50.8	51.1	49.0	**14.**	50.4	49.9	50.7
4.	50.6	51.1	49.0	**15.**	49.4	49.5	49.0
5.	49.6	50.5	50.9	**16.**	50.7	49.0	50.0
6.	51.3	52.0	50.3	**17.**	50.8	49.5	50.9
7.	49.7	50.5	50.3	**18.**	48.5	50.3	49.3
8.	51.8	50.3	50.0	**19.**	49.6	50.6	49.4
9.	48.6	50.5	50.7	**20.**	50.9	49.4	49.7
10.	49.6	49.8	50.5	**21.**	54.1	49.8	48.5
11.	49.9	50.7	49.8	**22.**	50.2	49.6	51.5

44. Una alternativa de la gráfica p para la fracción defectuosa es la *gráfica np para el número de piezas defectuosas.* Esta gráfica tiene límite superior de control $UCL = n\bar{p} + 3\sqrt{n\bar{p}(1-\bar{p})}$, $LCL = n\bar{p} - 3\sqrt{n\bar{p}(1-\bar{p})}$, y el *número* de piezas defectuosas de cada muestra se determina en la gráfica. Construya esa gráfica para los datos del ejemplo 16.6. ¿El uso de una gráfica *np* siempre da el mismo mensaje que el uso de una gráfica *p* (es decir, las dos gráficas son equivalentes)?

45. Observaciones de resistencia (ohms) para subgrupos de cierto tipo de registro dieron en resumen las siguientes cantidades:

i	n_i	$\bar{x}_i$	s_i	i	n_i	$\bar{x}_i$	s_i
1	4	430.0	22.5	11	4	445.2	27.3
2	4	418.2	20.6	12	4	430.1	22.2
3	3	435.5	25.1	13	4	427.2	24.0
4	4	427.6	22.3	14	4	439.6	23.3
5	4	444.0	21.5	15	3	415.9	31.2
6	3	431.4	28.9	16	4	419.8	27.5
7	4	420.8	25.4	17	3	447.0	19.8
8	4	431.4	24.0	18	4	434.4	23.7
9	4	428.7	21.2	19	4	422.2	25.1
10	4	440.1	25.8	20	4	425.7	24.4

Construya límites de control apropiados. [*Sugerencia:* Use $\bar{\bar{x}} = \sum n_i\bar{x}_i / \sum n_i$ y $s^2 = \sum(n_i - 1)s_i^2 / \sum(n_i - 1)$.]

46. Sea α el número entre 0 y 1 y defina una secuencia $W_1, W_2, W_3, \ldots$ por $W_0 = \mu$ y $W_t = \alpha\bar{X}_t + (1-\alpha)W_{t-1}$ para $t = 1, 2, \ldots$ Sustituyendo por W_{t-1} su representación en términos de $\bar{X}_{t-1}$ y W_{t-2}, luego sustituyendo por W_{t-2}, y así sucesivamente, da por resultado

$$W_t = \alpha\bar{X}_t + \alpha(1-\alpha)\bar{X}_{t-1} + \ldots + \alpha(1-\alpha)^{t-1}\bar{X}_1 + (1-\alpha)^t\mu$$

El hecho que W_t depende no sólo de $\bar{X}_t$ sino también de promedios para puntos de tiempo pasado, aunque con pesos (exponencialmente) decrecientes, sugiere que cambios en la media del proceso se reflejarán con más rapidez en las W_t que en las $\bar{X}_t$ individuales.

a. Demuestre que $E(W_t) = \mu$.

b. Sea $\sigma_t^2 = V(W_t)$, y demuestre que

$$\sigma_t^2 = \frac{\alpha[1-(1-\alpha)^{2t}]}{2-\alpha} \cdot \frac{\sigma^2}{n}$$

c. Una *gráfica de control de promedio móvil ponderada exponencialmente* grafica las W_t y utiliza límites de control $\mu_0 \pm 3\sigma_t$ (o bien, $\bar{\bar{x}}$ en lugar de μ_0). Construya una gráfica de este tipo para la información del ejemplo 16.9, usando $\mu_0 = 40$.

Bibliografía

Box, George, Soren Bisgaard, y Conrad Fung, "An Explanation and Critique of Taguchi's Contributions to Quality Engineering", *Quality and Reliability Engineering International*, 1988: 123-131.

Montgomery, Douglas C., *Introduction to Statistical Quality Control* (6a. ed.), Wiley, Nueva York, 2008. Ésta es una introducción exhaustiva a los numerosos aspectos del control de calidad a aproximadamente el mismo nivel que este libro.

Ryan, Thomas P., *Statistical Methods for Quality Improvement*, 2a. ed., Wiley, Nueva York, 2000. Capta muy bien el sabor moderno del control de calidad con mínimas demandas sobre los conocimientos previos del lector.

Vardeman, Stephen B. y J. Marcus Jobe, *Statistical Quality Assurance Methods for Engineers*, Wiley, New York, 1999. Incluye temas tradicionales de calidad y también material de diseño experimental relevante para temas relativos a calidad; informal y es una autoridad.

Tablas de apéndice

Tabla A.1 Distribución binomial acumulada

$$B(x; n, p) = \sum_{y=0}^{x} b(y; n, p)$$

a. $n = 5$

		p														
		0.01	**0.05**	**0.10**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.75**	**0.80**	**0.90**	**0.95**	**0.99**
	0	.951	.774	.590	.328	.237	.168	.078	.031	.010	.002	.001	.000	.000	.000	.000
	1	.999	.977	.919	.737	.633	.528	.337	.188	.087	.031	.016	.007	.000	.000	.000
x	2	1.000	.999	.991	.942	.896	.837	.683	.500	.317	.163	.104	.058	.009	.001	.000
	3	1.000	1.000	1.000	.993	.984	.969	.913	.812	.663	.472	.367	.263	.081	.023	.001
	4	1.000	1.000	1.000	1.000	.999	.998	.990	.969	.922	.832	.763	.672	.410	.226	.049

b. $n = 10$

		p														
		0.01	**0.05**	**0.10**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.75**	**0.80**	**0.90**	**0.95**	**0.99**
	0	.904	.599	.349	.107	.056	.028	.006	.001	.000	.000	.000	.000	.000	.000	.000
	1	.996	.914	.736	.376	.244	.149	.046	.011	.002	.000	.000	.000	.000	.000	.000
	2	1.000	.988	.930	.678	.526	.383	.167	.055	.012	.002	.000	.000	.000	.000	.000
	3	1.000	.999	.987	.879	.776	.650	.382	.172	.055	.011	.004	.001	.000	.000	.000
	4	1.000	1.000	.998	.967	.922	.850	.633	.377	.166	.047	.020	.006	.000	.000	.000
x	5	1.000	1.000	1.000	.994	.980	.953	.834	.623	.367	.150	.078	.033	.002	.000	.000
	6	1.000	1.000	1.000	.999	.996	.989	.945	.828	.618	.350	.224	.121	.013	.001	.000
	7	1.000	1.000	1.000	1.000	1.000	.998	.988	.945	.833	.617	.474	.322	.070	.012	.000
	8	1.000	1.000	1.000	1.000	1.000	1.000	.998	.989	.954	.851	.756	.624	.264	.086	.004
	9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.994	.972	.944	.893	.651	.401	.096

c. $n = 15$

		p														
		0.01	**0.05**	**0.10**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.75**	**0.80**	**0.90**	**0.95**	**0.99**
	0	.860	.463	.206	.035	.013	.005	.000	.000	.000	.000	.000	.000	.000	.000	.000
	1	.990	.829	.549	.167	.080	.035	.005	.000	.000	.000	.000	.000	.000	.000	.000
	2	1.000	.964	.816	.398	.236	.127	.027	.004	.000	.000	.000	.000	.000	.000	.000
	3	1.000	.995	.944	.648	.461	.297	.091	.018	.002	.000	.000	.000	.000	.000	.000
	4	1.000	.999	.987	.836	.686	.515	.217	.059	.009	.001	.000	.000	.000	.000	.000
	5	1.000	1.000	.998	.939	.852	.722	.403	.151	.034	.004	.001	.000	.000	.000	.000
	6	1.000	1.000	1.000	.982	.943	.869	.610	.304	.095	.015	.004	.001	.000	.000	.000
x	7	1.000	1.000	1.000	.996	.983	.950	.787	.500	.213	.050	.017	.004	.000	.000	.000
	8	1.000	1.000	1.000	.999	.996	.985	.905	.696	.390	.131	.057	.018	.000	.000	.000
	9	1.000	1.000	1.000	1.000	.999	.996	.966	.849	.597	.278	.148	.061	.002	.000	.000
	10	1.000	1.000	1.000	1.000	1.000	.999	.991	.941	.783	.485	.314	.164	.013	.001	.000
	11	1.000	1.000	1.000	1.000	1.000	1.000	.998	.982	.909	.703	.539	.352	.056	.005	.000
	12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.996	.973	.873	.764	.602	.184	.036	.000
	13	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.995	.965	.920	.833	.451	.171	.010
	14	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.995	.987	.965	.794	.537	.140

(*continúa*)

Tabla A.1 Distribución binomial acumulada *(continuación)*

$$B(x; n, p) = \sum_{y=0}^{x} b(y; n, p)$$

d. $n = 20$

		p														
		0.01	**0.05**	**0.10**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.75**	**0.80**	**0.90**	**0.95**	**0.99**
x	0	.818	.358	.122	.012	.003	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000
	1	.983	.736	.392	.069	.024	.008	.001	.000	.000	.000	.000	.000	.000	.000	.000
	2	.999	.925	.677	.206	.091	.035	.004	.000	.000	.000	.000	.000	.000	.000	.000
	3	1.000	.984	.867	.411	.225	.107	.016	.001	.000	.000	.000	.000	.000	.000	.000
	4	1.000	.997	.957	.630	.415	.238	.051	.006	.000	.000	.000	.000	.000	.000	.000
	5	1.000	1.000	.989	.804	.617	.416	.126	.021	.002	.000	.000	.000	.000	.000	.000
	6	1.000	1.000	.998	.913	.786	.608	.250	.058	.006	.000	.000	.000	.000	.000	.000
	7	1.000	1.000	1.000	.968	.898	.772	.416	.132	.021	.001	.000	.000	.000	.000	.000
	8	1.000	1.000	1.000	.990	.959	.887	.596	.252	.057	.005	.001	.000	.000	.000	.000
	9	1.000	1.000	1.000	.997	.986	.952	.755	.412	.128	.017	.004	.001	.000	.000	.000
	10	1.000	1.000	1.000	.999	.996	.983	.872	.588	.245	.048	.014	.003	.000	.000	.000
	11	1.000	1.000	1.000	1.000	.999	.995	.943	.748	.404	.113	.041	.010	.000	.000	.000
	12	1.000	1.000	1.000	1.000	1.000	.999	.979	.868	.584	.228	.102	.032	.000	.000	.000
	13	1.000	1.000	1.000	1.000	1.000	1.000	.994	.942	.750	.392	.214	.087	.002	.000	.000
	14	1.000	1.000	1.000	1.000	1.000	1.000	.998	.979	.874	.584	.383	.196	.011	.000	.000
	15	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.994	.949	.762	.585	.370	.043	.003	.000
	16	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.984	.893	.775	.589	.133	.016	.000
	17	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.996	.965	.909	.794	.323	.075	.001
	18	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.992	.976	.931	.608	.264	.017
	19	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.997	.988	.878	.642	.182

(*continúa*)

Tabla A.1 Distribución binomial acumulada *(continuación)*

$$B(x; n, p) = \sum_{y=0}^{x} b(y; n, p)$$

e. $n = 25$

		p														
		0.01	**0.05**	**0.10**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.75**	**0.80**	**0.90**	**0.95**	**0.99**
	0	.778	.277	.072	.004	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	1	.974	.642	.271	.027	.007	.002	.000	.000	.000	.000	.000	.000	.000	.000	.000
	2	.998	.873	.537	.098	.032	.009	.000	.000	.000	.000	.000	.000	.000	.000	.000
	3	1.000	.966	.764	.234	.096	.033	.002	.000	.000	.000	.000	.000	.000	.000	.000
	4	1.000	.993	.902	.421	.214	.090	.009	.000	.000	.000	.000	.000	.000	.000	.000
	5	1.000	.999	.967	.617	.378	.193	.029	.002	.000	.000	.000	.000	.000	.000	.000
	6	1.000	1.000	.991	.780	.561	.341	.074	.007	.000	.000	.000	.000	.000	.000	.000
	7	1.000	1.000	.998	.891	.727	.512	.154	.022	.001	.000	.000	.000	.000	.000	.000
	8	1.000	1.000	1.000	.953	.851	.677	.274	.054	.004	.000	.000	.000	.000	.000	.000
	9	1.000	1.000	1.000	.983	.929	.811	.425	.115	.013	.000	.000	.000	.000	.000	.000
	10	1.000	1.000	1.000	.994	.970	.902	.586	.212	.034	.002	.000	.000	.000	.000	.000
	11	1.000	1.000	1.000	.998	.980	.956	.732	.345	.078	.006	.001	.000	.000	.000	.000
x	12	1.000	1.000	1.000	1.000	.997	.983	.846	.500	.154	.017	.003	.000	.000	.000	.000
	13	1.000	1.000	1.000	1.000	.999	.994	.922	.655	.268	.044	.020	.002	.000	.000	.000
	14	1.000	1.000	1.000	1.000	1.000	.998	.966	.788	.414	.098	.030	.006	.000	.000	.000
	15	1.000	1.000	1.000	1.000	1.000	1.000	.987	.885	.575	.189	.071	.017	.000	.000	.000
	16	1.000	1.000	1.000	1.000	1.000	1.000	.996	.946	.726	.323	.149	.047	.000	.000	.000
	17	1.000	1.000	1.000	1.000	1.000	1.000	.999	.978	.846	.488	.273	.109	.002	.000	.000
	18	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.993	.926	.659	.439	.220	.009	.000	.000
	19	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.998	.971	.807	.622	.383	.033	.001	.000
	20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.991	.910	.786	.579	.098	.007	.000
	21	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.998	.967	.904	.766	.236	.034	.000
	22	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.991	.968	.902	.463	.127	.002
	23	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.998	.993	.973	.729	.358	.026
	24	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.996	.928	.723	.222

Tabla A.2 Distribución acumulada de Poisson

$$F(x; \mu) = \sum_{y=0}^{x} \frac{e^{-\mu}\mu^{y}}{y!}$$

		.1	**.2**	**.3**	**.4**	**.5**	**.6**	**.7**	**.8**	**.9**	**1.0**
	0	.905	.819	.741	.670	.607	.549	.497	.449	.407	.368
	1	.995	.982	.963	.938	.910	.878	.844	.809	.772	.736
	2	1.000	.999	.996	.992	.986	.977	.966	.953	.937	.920
x	3		1.000	1.000	.999	.998	.997	.994	.991	.987	.981
	4				1.000	1.000	1.000	.999	.999	.998	.996
	5							1.000	1.000	1.000	.999
	6										1.000

(*continúa*)

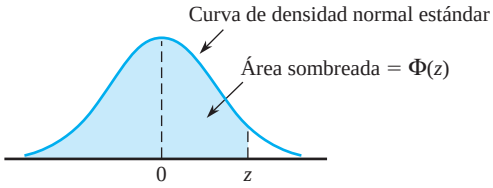
Tabla A.2 Distribución acumulada de Poisson *(continuación)*

$$F(x;\mu) = \sum_{y=0}^{x} \frac{e^{-\mu}\mu^{y}}{y!}$$

	x	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0	15.0	20.0
	0	.135	.050	.018	.007	.002	.001	.000	.000	.000	.000	.000
	1	.406	.199	.092	.040	.017	.007	.003	.001	.000	.000	.000
	2	.677	.423	.238	.125	.062	.030	.014	.006	.003	.000	.000
	3	.857	.647	.433	.265	.151	.082	.042	.021	.010	.000	.000
	4	.947	.815	.629	.440	.285	.173	.100	.055	.029	.001	.000
	5	.983	.916	.785	.616	.446	.301	.191	.116	.067	.003	.000
	6	.995	.966	.889	.762	.606	.450	.313	.207	.130	.008	.000
	7	.999	.988	.949	.867	.744	.599	.453	.324	.220	.018	.001
	8	1.000	.996	.979	.932	.847	.729	.593	.456	.333	.037	.002
	9		.999	.992	.968	.916	.830	.717	.587	.458	.070	.005
	10		1.000	.997	.986	.957	.901	.816	.706	.583	.118	.011
	11			.999	.995	.980	.947	.888	.803	.697	.185	.021
	12			1.000	.998	.991	.973	.936	.876	.792	.268	.039
	13				.999	.996	.987	.966	.926	.864	.363	.066
	14				1.000	.999	.994	.983	.959	.917	.466	.105
	15					.999	.998	.992	.978	.951	.568	.157
	16					1.000	.999	.996	.989	.973	.664	.221
	17						1.000	.998	.995	.986	.749	.297
	18							.999	.998	.993	.819	.381
	19							1.000	.999	.997	.875	.470
	20								1.000	.998	.917	.559
	21									.999	.947	.644
	22									1.000	.967	.721
	23										.981	.787
	24										.989	.843
	25										.994	.888
	26										.997	.922
	27										.998	.948
	28										.999	.966
	29										1.000	.978
	30											.987
	31											.992
	32											.995
	33											.997
	34											.999
	35											.999
	36											1.000

Tabla A.3 Áreas de la curva normal estándar

$\Phi(z) = P(Z \le z)$



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0038
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0352	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3482
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

(*continúa*)

Tabla A.3 Áreas de la curva normal estándar *(continuación)*

$\Phi(z) = P(Z \leq z)$

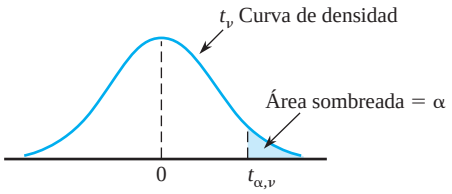
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9278	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Tabla A.4 La función gamma incompleta

$$F(x;\alpha) = \int_0^x \frac{1}{\Gamma(\alpha)} y^{\alpha-1} e^{-y}\, dy$$

x	1	2	3	4	5	6	7	8	9	10
1	.632	.264	.080	.019	.004	.001	.000	.000	.000	.000
2	.865	.594	.323	.143	.053	.017	.005	.001	.000	.000
3	.950	.801	.577	.353	.185	.084	.034	.012	.004	.001
4	.982	.908	.762	.567	.371	.215	.111	.051	.021	.008
5	.993	.960	.875	.735	.560	.384	.238	.133	.068	.032
6	.998	.983	.938	.849	.715	.554	.394	.256	.153	.084
7	.999	.993	.970	.918	.827	.699	.550	.401	.271	.170
8	1.000	.997	.986	.958	.900	.809	.687	.547	.407	.283
9		.999	.994	.979	.945	.884	.793	.676	.544	.413
10		1.000	.997	.990	.971	.933	.870	.780	.667	.542
11			.999	.995	.985	.962	.921	.857	.768	.659
12			1.000	.998	.992	.980	.954	.911	.845	.758
13				.999	.996	.989	.974	.946	.900	.834
14				1.000	.998	.994	.986	.968	.938	.891
15					.999	.997	.992	.982	.963	.930

Tabla A.5 Valores críticos para distribuciones *t*

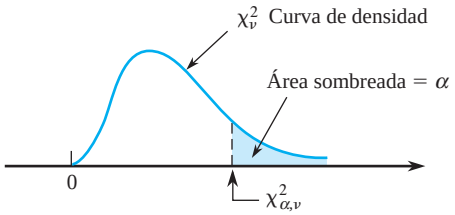


v	.10	.05	.025	.01	.005	.001	.0005
1	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.767
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
32	1.309	1.694	2.037	2.449	2.738	3.365	3.622
34	1.307	1.691	2.032	2.441	2.728	3.348	3.601
36	1.306	1.688	2.028	2.434	2.719	3.333	3.582
38	1.304	1.686	2.024	2.429	2.712	3.319	3.566
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
50	1.299	1.676	2.009	2.403	2.678	3.262	3.496
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

Tabla A.6 Valores críticos de tolerancia para distribuciones normales de población

	Intervalos bilaterales						Intervalos unilaterales					
Nivel de confianza	**95%**			**99%**			**95%**			**99%**		
% de población capturada	**≥ 90%**	**≥ 95%**	**≥ 99%**	**≥ 90%**	**≥ 95%**	**≥ 99%**	**≥ 90%**	**≥ 95%**	**≥ 99%**	**≥ 90%**	**≥ 95%**	**≥ 99%**
2	32.019	37.674	48.430	160.193	188.491	242.300	20.581	26.260	37.094	103.029	131.426	185.617
3	8.380	9.916	12.861	18.930	22.401	29.055	6.156	7.656	10.553	13.995	17.370	23.896
4	5.369	6.370	8.299	9.398	11.150	14.527	4.162	5.144	7.042	7.380	9.083	12.387
5	4.275	5.079	6.634	6.612	7.855	10.260	3.407	4.203	5.741	5.362	6.578	8.939
6	3.712	4.414	5.775	5.337	6.345	8.301	3.006	3.708	5.062	4.411	5.406	7.335
7	3.369	4.007	5.248	4.613	5.488	7.187	2.756	3.400	4.642	3.859	4.728	6.412
8	3.136	3.732	4.891	4.147	4.936	6.468	2.582	3.187	4.354	3.497	4.285	5.812
9	2.967	3.532	4.631	3.822	4.550	5.966	2.454	3.031	4.143	3.241	3.972	5.389
10	2.839	3.379	4.433	3.582	4.265	5.594	2.355	2.911	3.981	3.048	3.738	5.074
11	2.737	3.259	4.277	3.397	4.045	5.308	2.275	2.815	3.852	2.898	3.556	4.829
12	2.655	3.162	4.150	3.250	3.870	5.079	2.210	2.736	3.747	2.777	3.410	4.633
13	2.587	3.081	4.044	3.130	3.727	4.893	2.155	2.671	3.659	2.677	3.290	4.472
14	2.529	3.012	3.955	3.029	3.608	4.737	2.109	2.615	3.585	2.593	3.189	4.337
15	2.480	2.954	3.878	2.945	3.507	4.605	2.068	2.566	3.520	2.522	3.102	4.222
16	2.437	2.903	3.812	2.872	3.421	4.492	2.033	2.524	3.464	2.460	3.028	4.123
Tamaño de la muestra *n* **17**	2.400	2.858	3.754	2.808	3.345	4.393	2.002	2.486	3.414	2.405	2.963	4.037
18	2.366	2.819	3.702	2.753	3.279	4.307	1.974	2.453	3.370	2.357	2.905	3.960
19	2.337	2.784	3.656	2.703	3.221	4.230	1.949	2.423	3.331	2.314	2.854	3.892
20	2.310	2.752	3.615	2.659	3.168	4.161	1.926	2.396	3.295	2.276	2.808	3.832
25	2.208	2.631	3.457	2.494	2.972	3.904	1.838	2.292	3.158	2.129	2.633	3.601
30	2.140	2.549	3.350	2.385	2.841	3.733	1.777	2.220	3.064	2.030	2.516	3.447
35	2.090	2.490	3.272	2.306	2.748	3.611	1.732	2.167	2.995	1.957	2.430	3.334
40	2.052	2.445	3.213	2.247	2.677	3.518	1.697	2.126	2.941	1.902	2.364	3.249
45	2.021	2.408	3.165	2.200	2.621	3.444	1.669	2.092	2.898	1.857	2.312	3.180
50	1.996	2.379	3.126	2.162	2.576	3.385	1.646	2.065	2.863	1.821	2.269	3.125
60	1.958	2.333	3.066	2.103	2.506	3.293	1.609	2.022	2.807	1.764	2.202	3.038
70	1.929	2.299	3.021	2.060	2.454	3.225	1.581	1.990	2.765	1.722	2.153	2.974
80	1.907	2.272	2.986	2.026	2.414	3.173	1.559	1.965	2.733	1.688	2.114	2.924
90	1.889	2.251	2.958	1.999	2.382	3.130	1.542	1.944	2.706	1.661	2.082	2.883
100	1.874	2.233	2.934	1.977	2.355	3.096	1.527	1.927	2.684	1.639	2.056	2.850
150	1.825	2.175	2.859	1.905	2.270	2.983	1.478	1.870	2.611	1.566	1.971	2.741
200	1.798	2.143	2.816	1.865	2.222	2.921	1.450	1.837	2.570	1.524	1.923	2.679
250	1.780	2.121	2.788	1.839	2.191	2.880	1.431	1.815	2.542	1.496	1.891	2.638
300	1.767	2.106	2.767	1.820	2.169	2.850	1.417	1.800	2.522	1.476	1.868	2.608
∞	1.645	1.960	2.576	1.645	1.960	2.576	1.282	1.645	2.326	1.282	1.645	2.326

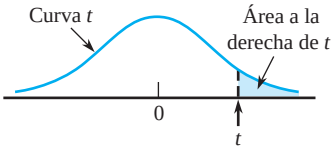
Tabla A.7 Valores críticos para distribuciones chi-cuadrada



	.995	.99	.975	.95	.90	.10	.05	.025	.01	.005
1	0.000	0.000	0.001	0.004	0.016	2.706	3.843	5.025	6.637	7.882
2	0.010	0.020	0.051	0.103	0.211	4.605	5.992	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.344	12.837
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.832	15.085	16.748
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.440	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.012	18.474	20.276
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.534	20.090	21.954
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.022	21.665	23.587
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.724	26.755
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.041	19.812	22.362	24.735	27.687	29.817
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.600	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.577	32.799
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.407	7.564	8.682	10.085	24.769	27.587	30.190	33.408	35.716
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.843	7.632	8.906	10.117	11.651	27.203	30.143	32.852	36.190	38.580
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.033	8.897	10.283	11.591	13.240	29.615	32.670	35.478	38.930	41.399
22	8.643	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.260	10.195	11.688	13.090	14.848	32.007	35.172	38.075	41.637	44.179
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.558
25	10.519	11.523	13.120	14.611	16.473	34.381	37.652	40.646	44.313	46.925
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.807	12.878	14.573	16.151	18.114	36.741	40.113	43.194	46.962	49.642
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.120	14.256	16.147	17.708	19.768	39.087	42.557	45.772	49.586	52.333
30	13.787	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
31	14.457	15.655	17.538	19.280	21.433	41.422	44.985	48.231	52.190	55.000
32	15.134	16.362	18.291	20.072	22.271	42.585	46.194	49.480	53.486	56.328
33	15.814	17.073	19.046	20.866	23.110	43.745	47.400	50.724	54.774	57.646
34	16.501	17.789	19.806	21.664	23.952	44.903	48.602	51.966	56.061	58.964
35	17.191	18.508	20.569	22.465	24.796	46.059	49.802	53.203	57.340	60.272
36	17.887	19.233	21.336	23.269	25.643	47.212	50.998	54.437	58.619	61.581
37	18.584	19.960	22.105	24.075	26.492	48.363	52.192	55.667	59.891	62.880
38	19.289	20.691	22.878	24.884	27.343	49.513	53.384	56.896	61.162	64.181
39	19.994	21.425	23.654	25.695	28.196	50.660	54.572	58.119	62.426	65.473
40	20.706	22.164	24.433	26.509	29.050	51.805	55.758	59.342	63.691	66.766

Para $\nu > 40$, $\chi^2_{a,\nu} \approx \nu\left(1 - \frac{2}{9\nu} + z_a\sqrt{\frac{2}{9\nu}}\right)^3$

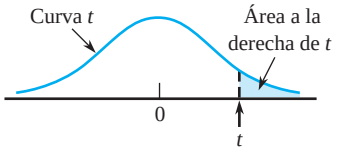
Tabla A.8 Áreas de cola de la curva *t*



t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
0.0	.500	.500	.500	.500	.500	.500	.500	.500	.500	.500	.500	.500	.500	.500	.500	.500	.500	.500
0.1	.468	.465	.463	.463	.462.	.462	.462	.461	.461	.461	.461	.461	.461	.461	.461	.461	.461	.461
0.2	.437	.430	.427	.426	.425	.424	.424	.423	.423	.423	.423	.422	.422	.422	.422	.422	.422	.422
0.3	.407	.396	.392	.390	.388	.387	.386	.386	.386	.385	.385	.385	.384	.384	.384	.384	.384	.384
0.4	.379	.364	.358	.355	.353	.352	.351	.350	.349	.349	.348	.348	.348	.347	.347	.347	.347	.347
0.5	.352	.333	.326	.322	.319	.317	.316	.315	.315	.314	.313	.313	.313	.312	.312	.312	.312	.312
0.6	.328	.305	.295	.290	.287	.285	.284	.283	.282	.281	.280	.280	.279	.279	.279	.278	.278	.278
0.7	.306	.278	.267	.261	.258	.255	.253	.252	.251	.250	.249	.249	.248	.247	.247	.247	.247	.246
0.8	.285	.254	.241	.234	.230	.227	.225	.223	.222	.221	.220	.220	.219	.218	.218	.218	.217	.217
0.9	.267	.232	.217	.210	.205	.201	.199	.197	.196	.195	.194	.193	.192	.191	.191	.191	.190	.190
1.0	.250	.211	.196	.187	.182	.178	.175	.173	.172	.170	.169	.169	.168	.167	.167	.166	.166	.165
1.1	.235	.193	.176	.167	.162	.157	.154	.152	.150	.149	.147	.146	.146	.144	.144	.144	.143	.143
1.2	.221	.177	.158	.148	.142	.138	.135	.132	.130	.129	.128	.127	.126	.124	.124	.124	.123	.123
1.3	.209	.162	.142	.132	.125	.121	.117	.115	.113	.111	.110	.109	.108	.107	.107	.106	.105	.105
1.4	.197	.148	.128	.117	.110	.106	.102	.100	.098	.096	.095	.093	.092	.091	.091	.090	.090	.089
1.5	.187	.136	.115	.104	.097	.092	.089	.086	.084	.082	.081	.080	.079	.077	.077	.077	.076	.075
1.6	.178	.125	.104	.092	.085	.080	.077	.074	.072	.070	.069	.068	.067	.065	.065	.065	.064	.064
1.7	.169	.116	.094	.082	.075	.070	.065	.064	.062	.060	.059	.057	.056	.055	.055	.054	.054	.053
1.8	.161	.107	.085	.073	.066	.061	.057	.055	.053	.051	.050	.049	.048	.046	.046	.045	.045	.044
1.9	.154	.099	.077	.065	.058	.053	.050	.047	.045	.043	.042	.041	.040	.038	.038	.038	.037	.037
2.0	.148	.092	.070	.058	.051	.046	.043	.040	.038	.037	.035	.034	.033	.032	.032	.031	.031	.030
2.1	.141	.085	.063	.052	.045	.040	.037	.034	.033	.031	.030	.029	.028	.027	.027	.026	.025	.025
2.2	.136	.079	.058	.046	.040	.035	.032	.029	.028	.026	.025	.024	.023	.022	.022	.021	.021	.021
2.3	.131	.074	.052	.041	.035	.031	.027	.025	.023	.022	.021	.020	.019	.018	.018	.018	.017	.017
2.4	.126	.069	.048	.037	.031	.027	.024	.022	.020	.019	.018	.017	.016	.015	.015	.014	.014	.014
2.5	.121	.065	.044	.033	.027	.023	.020	.018	.017	.016	.015	.014	.013	.012	.012	.012	.011	.011
2.6	.117	.061	.040	.030	.024	.020	.018	.016	.014	.013	.012	.012	.011	.010	.010	.010	.009	.009
2.7	.113	.057	.037	.027	.021	.018	.015	.014	.012	.011	.010	.010	.009	.008	.008	.008	.008	.007
2.8	.109	.054	.034	.024	.019	.016	.013	.012	.010	.009	.009	.008	.008	.007	.007	.006	.006	.006
2.9	.106	.051	.031	.022	.017	.014	.011	.010	.009	.008	.007	.007	.006	.005	.005	.005	.005	.005
3.0	.102	.048	.029	.020	.015	.012	.010	.009	.007	.007	.006	.006	.005	.004	.004	.004	.004	.004
3.1	.099	.045	.027	.018	.013	.011	.009	.007	.006	.006	.005	.005	.004	.004	.004	.003	.003	.003
3.2	.096	.043	.025	.016	.012	.009	.008	.006	.005	.005	.004	.004	.003	.003	.003	.003	.003	.002
3.3	.094	.040	.023	.015	.011	.008	.007	.005	.005	.004	.004	.003	.003	.002	.002	.002	.002	.002
3.4	.091	.038	.021	.014	.010	.007	.006	.005	.004	.003	.003	.003	.002	.002	.002	.002	.002	.002
3.5	.089	.036	.020	.012	.009	.006	.005	.004	.003	.003	.002	.002	.002	.002	.002	.001	.001	.001
3.6	.086	.035	.018	.011	.008	.006	.004	.004	.003	.002	.002	.002	.002	.001	.001	.001	.001	.001
3.7	.084	.033	.017	.010	.007	.005	.004	.003	.002	.002	.002	.002	.001	.001	.001	.001	.001	.001
3.8	.082	.031	.016	.010	.006	.004	.003	.003	.002	.002	.001	.001	.001	.001	.001	.001	.001	.001
3.9	.080	.030	.015	.009	.006	.004	.003	.002	.002	.001	.001	.001	.001	.001	.001	.001	.001	.001
4.0	.078	.029	.014	.008	.005	.004	.003	.002	.002	.001	.001	.001	.001	.001	.001	.001	.000	.000

(*continúa*)

Tabla A.8 Áreas de cola de la curva *t (continuación)*



t	19	20	21	22	23	24	25	26	27	28	29	30	35	40	60	120	∞(=z)
0.0	.500	.500	.500	.500	.500	.500	.500	.500	.500	.500	.500	.500	.500	.500	.500	.500	.500
0.1	.461	.461	.461	.461	.461	.461	.461	.461	.461	.461	.461	.461	.460	.460	.460	.460	.460
0.2	.422	.422	.422	.422	.422	.422	.422	.422	.421	.421	.421	.421	.421	.421	.421	.421	.421
0.3	.384	.384	.384	.383	.383	.383	.383	.383	.383	.383	.383	.383	.383	.383	.383	.382	.382
0.4	.347	.347	.347	.347	.346	.346	.346	.346	.346	.346	.346	.346	.346	.346	.345	.345	.345
0.5	.311	.311	.311	.311	.311	.311	.311	.311	.311	.310	.310	.310	.310	.310	.309	.309	.309
0.6	.278	.278	.278	.277	.277	.277	.277	.277	.277	.277	.277	.277	.276	.276	.275	.275	.274
0.7	.246	.246	.246	.246	.245	.245	.245	.245	.245	.245	.245	.245	.244	.244	.243	.243	.242
0.8	.217	.217	.216	.216	.216	.216	.216	.215	.215	.215	.215	.215	.215	.214	.213	.213	.212
0.9	.190	.189	.189	.189	.189	.189	.188	.188	.188	.188	.188	.188	.187	.187	.186	.185	.184
1.0	.165	.165	.164	.164	.164	.164	.163	.163	.163	.163	.163	.163	.162	.162	.161	.160	.159
1.1	.143	.142	.142	.142	.141	.141	.141	.141	.141	.140	.140	.140	.139	.139	.138	.137	.136
1.2	.122	.122	.122	.121	.121	.121	.121	.120	.120	.120	.120	.120	.119	.119	.117	.116	.115
1.3	.105	.104	.104	.104	.103	.103	.103	.103	.102	.102	.102	.102	.101	.101	.099	.098	.097
1.4	.089	.089	.088	.088	.087	.087	.087	.087	.086	.086	.086	.086	.085	.085	.083	.082	.081
1.5	.075	.075	.074	.074	.074	.073	.073	.073	.073	.072	.072	.072	.071	.071	.069	.068	.067
1.6	.063	.063	.062	.062	.062	.061	.061	.061	.061	.060	.060	.060	.059	.059	.057	.056	.055
1.7	.053	.052	.052	.052	.051	.051	.051	.051	.050	.050	.050	.050	.049	.048	.047	.046	.045
1.8	.044	.043	.043	.043	.042	.042	.042	.042	.042	.041	.041	.041	.040	.040	.038	.037	.036
1.9	.036	.036	.036	.035	.035	.035	.035	.034	.034	.034	.034	.034	.033	.032	.031	.030	.029
2.0	.030	.030	.029	.029	.029	.028	.028	.028	.028	.028	.027	.027	.027	.026	.025	.024	.023
2.1	.025	.024	.024	.024	.023	.023	.023	.023	.023	.022	.022	.022	.022	.021	.020	.019	.018
2.2	.020	.020	.020	.019	.019	.019	.019	.018	.018	.018	.018	.018	.017	.017	.016	.015	.014
2.3	.016	.016	.016	.016	.015	.015	.015	.015	.015	.015	.014	.014	.014	.013	.012	.012	.011
2.4	.013	.013	.013	.013	.012	.012	.012	.012	.012	.012	.012	.011	.011	.011	.010	.009	.008
2.5	.011	.011	.010	.010	.010	.010	.010	.010	.009	.009	.009	.009	.009	.008	.008	.007	.006
2.6	.009	.009	.008	.008	.008	.008	.008	.008	.007	.007	.007	.007	.007	.007	.006	.005	.005
2.7	.007	.007	.007	.007	.006	.006	.006	.006	.006	.006	.006	.006	.005	.005	.004	.004	.003
2.8	.006	.006	.005	.005	.005	.005	.005	.005	.005	.005	.005	.004	.004	.004	.003	.003	.003
2.9	.005	.004	.004	.004	.004	.004	.004	.004	.004	.004	.004	.003	.003	.003	.003	.002	.002
3.0	.004	.004	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003	.002	.002	.002	.002	.001
3.1	.003	.003	.003	.003	.003	.002	.002	.002	.002	.002	.002	.002	.002	.002	.001	.001	.001
3.2	.002	.002	.002	.002	.002	.002	.002	.002	.002	.002	.002	.002	.001	.001	.001	.001	.001
3.3	.002	.002	.002	.002	.002	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.000
3.4	.002	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.000	.000
3.5	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.000	.000	.000
3.6	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.000	.000	.000	.000	.000
3.7	.001	.001	.001	.001	.001	.001	.001	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000
3.8	.001	.001	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
3.9	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
4.0	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000

Tabla A.9 Valores críticos de la distribución *F*

ν_2 = denominador gl		ν_1 = numerador gl								
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	**.100**	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86
	.050	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
	.010	4052.20	4999.50	5403.40	5624.60	5763.60	5859.00	5928.40	5981.10	6022.50
	.001	405,284	500,000	540,379	562,500	576,405	585,937	592,873	598,144	602,284
2	**.100**	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
	.050	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
	.010	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
	.001	998.50	999.00	999.17	999.25	999.30	999.33	999.36	999.37	999.39
3	**.100**	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
	.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
	.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
	.001	167.03	148.50	141.11	137.10	134.58	132.85	131.58	130.62	129.86
4	**.100**	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
	.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
	.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
	.001	74.14	61.25	56.18	53.44	51.71	50.53	49.66	49.00	48.47
5	**.100**	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32
	.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
	.010	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
	.001	47.18	37.12	33.20	31.09	29.75	28.83	28.16	27.65	27.24
6	**.100**	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96
	.050	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
	.010	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
	.001	35.51	27.00	23.70	21.92	20.80	20.03	19.46	19.03	18.69
7	**.100**	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72
	.050	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
	.010	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
	.001	29.25	21.69	18.77	17.20	16.21	15.52	15.02	14.63	14.33
8	**.100**	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56
	.050	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
	.010	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
	.001	25.41	18.49	15.83	14.39	13.48	12.86	12.40	12.05	11.77
9	**.100**	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44
	.050	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
	.010	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
	.001	22.86	16.39	13.90	12.56	11.71	11.13	10.70	10.37	10.11
10	**.100**	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35
	.050	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
	.010	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
	.001	21.04	14.91	12.55	11.28	10.48	9.93	9.52	9.20	8.96
11	**.100**	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27
	.050	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
	.010	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
	.001	19.69	13.81	11.56	10.35	9.58	9.05	8.66	8.35	8.12
12	**.100**	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21
	.050	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
	.010	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
	.001	18.64	12.97	10.80	9.63	8.89	8.38	8.00	7.71	7.48

(*continúa*)

Tabla A.9 Valores críticos de la distribución *F (continuación)*

ν_1 = numerador gl

10	12	15	20	25	30	40	50	60	120	1000
60.19	60.71	61.22	61.74	62.05	62.26	62.53	62.69	62.79	63.06	63.30
241.88	243.91	245.95	248.01	249.26	250.10	251.14	251.77	252.20	253.25	254.19
6055.80	6106.30	6157.30	6208.70	6239.80	6260.60	6286.80	6302.50	6313.00	6339.40	6362.70
605,621	610,668	615,764	620,908	624,017	626,099	628,712	630,285	631,337	633,972	636,301
9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.47	9.48	9.49
19.40	19.41	19.43	19.45	19.46	19.46	19.47	19.48	19.48	19.49	19.49
99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.48	99.49	99.50
999.40	999.42	999.43	999.45	999.46	999.47	999.47	999.48	999.48	999.49	999.50
5.23	5.22	5.20	5.18	5.17	5.17	5.16	5.15	5.15	5.14	5.13
8.79	8.74	8.70	8.66	8.63	8.62	8.59	8.58	8.57	8.55	8.53
27.23	27.05	26.87	26.69	26.58	26.50	26.41	26.35	26.32	26.22	26.14
129.25	128.32	127.37	126.42	125.84	125.45	124.96	124.66	124.47	123.97	123.53
3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.80	3.79	3.78	3.76
5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.70	5.69	5.66	5.63
14.55	14.37	14.20	14.02	13.91	13.84	13.75	13.69	13.65	13.56	13.47
48.05	47.41	46.76	46.10	45.70	45.43	45.09	44.88	44.75	44.40	44.09
3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.15	3.14	3.12	3.11
4.74	4.68	4.62	4.56	4.52	4.50	4.46	4.44	4.43	4.40	4.37
10.05	9.89	9.72	9.55	9.45	9.38	9.29	9.24	9.20	9.11	9.03
26.92	26.42	25.91	25.39	25.08	24.87	24.60	24.44	24.33	24.06	23.82
2.94	2.90	2.87	2.84	2.81	2.80	2.78	2.77	2.76	2.74	2.72
4.06	4.00	3.94	3.87	3.83	3.81	3.77	3.75	3.74	3.70	3.67
7.87	7.72	7.56	7.40	7.30	7.23	7.14	7.09	7.06	6.97	6.89
18.41	17.99	17.56	17.12	16.85	16.67	16.44	16.31	16.21	15.98	15.77
2.70	2.67	2.63	2.59	2.57	2.56	2.54	2.52	2.51	2.49	2.47
3.64	3.57	3.51	3.44	3.40	3.38	3.34	3.32	3.30	3.27	3.23
6.62	6.47	6.31	6.16	6.06	5.99	5.91	5.86	5.82	5.74	5.66
14.08	13.71	13.32	12.93	12.69	12.53	12.33	12.20	12.12	11.91	11.72
2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.35	2.34	2.32	2.30
3.35	3.28	3.22	3.15	3.11	3.08	3.04	3.02	3.01	2.97	2.93
5.81	5.67	5.52	5.36	5.26	5.20	5.12	5.07	5.03	4.95	4.87
11.54	11.19	10.84	10.48	10.26	10.11	9.92	9.80	9.73	9.53	9.36
2.42	2.38	2.34	2.30	2.27	2.25	2.23	2.22	2.21	2.18	2.16
3.14	3.07	3.01	2.94	2.89	2.86	2.83	2.80	2.79	2.75	2.71
5.26	5.11	4.96	4.81	4.71	4.65	4.57	4.52	4.48	4.40	4.32
9.89	9.57	9.24	8.90	8.69	8.55	8.37	8.26	8.19	8.00	7.84
2.32	2.28	2.24	2.20	2.17	2.16	2.13	2.12	2.11	2.08	2.06
2.98	2.91	2.85	2.77	2.73	2.70	2.66	2.64	2.62	2.58	2.54
4.85	4.71	4.56	4.41	4.31	4.25	4.17	4.12	4.08	4.00	3.92
8.75	8.45	8.13	7.80	7.60	7.47	7.30	7.19	7.12	6.94	6.78
2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.04	2.03	2.00	1.98
2.85	2.79	2.72	2.65	2.60	2.57	2.53	2.51	2.49	2.45	2.41
4.54	4.40	4.25	4.10	4.01	3.94	3.86	3.81	3.78	3.69	3.61
7.92	7.63	7.32	7.01	6.81	6.68	6.52	6.42	6.35	6.18	6.02
2.19	2.15	2.10	2.06	2.03	2.01	1.99	1.97	1.96	1.93	1.91
2.75	2.69	2.62	2.54	2.50	2.47	2.43	2.40	2.38	2.34	2.30
4.30	4.16	4.01	3.86	3.76	3.70	3.62	3.57	3.54	3.45	3.37
7.29	7.00	6.71	6.40	6.22	6.09	5.93	5.83	5.76	5.59	5.44

(*continúa*)

Tabla A.9 Valores críticos de la distribución *F (continuación)*

ν_2 = denominador gl		ν_1 = numerador gl								
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
13	**.100**	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16
	.050	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
	.010	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
	.001	17.82	12.31	10.21	9.07	8.35	7.86	7.49	7.21	6.98
14	**.100**	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12
	.050	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
	.010	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
	.001	17.14	11.78	9.73	8.62	7.92	7.44	7.08	6.80	6.58
15	**.100**	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09
	.050	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
	.010	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
	.001	16.59	11.34	9.34	8.25	7.57	7.09	6.74	6.47	6.26
16	**.100**	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06
	.050	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
	.010	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
	.001	16.12	10.97	9.01	7.94	7.27	6.80	6.46	6.19	5.98
17	**.100**	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03
	.050	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
	.010	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68
	.001	15.72	10.66	8.73	7.68	7.02	6.56	6.22	5.96	5.75
18	**.100**	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00
	.050	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
	.010	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
	.001	15.38	10.39	8.49	7.46	6.81	6.35	6.02	5.76	5.56
19	**.100**	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98
	.050	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
	.010	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
	.001	15.08	10.16	8.28	7.27	6.62	6.18	5.85	5.59	5.39
20	**.100**	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96
	.050	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
	.010	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
	.001	14.82	9.95	8.10	7.10	6.46	6.02	5.69	5.44	5.24
21	**.100**	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95
	.050	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
	.010	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
	.001	14.59	9.77	7.94	6.95	6.32	5.88	5.56	5.31	5.11
22	**.100**	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93
	.050	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
	.010	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
	.001	14.38	9.61	7.80	6.81	6.19	5.76	5.44	5.19	4.99
23	**.100**	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92
	.050	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
	.010	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
	.001	14.20	9.47	7.67	6.70	6.08	5.65	5.33	5.09	4.89
24	**.100**	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91
	.050	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
	.010	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
	.001	14.03	9.34	7.55	6.59	5.98	5.55	5.23	4.99	4.80

(*continúa*)

Tabla A.9 Valores críticos de la distribución *F (continuación)*

ν_1 = numerador gl

10	12	15	20	25	30	40	50	60	120	1000
2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.92	1.90	1.88	1.85
2.67	2.60	2.53	2.46	2.41	2.38	2.34	2.31	2.30	2.25	2.21
4.10	3.96	3.82	3.66	3.57	3.51	3.43	3.38	3.34	3.25	3.18
6.80	6.52	6.23	5.93	5.75	5.63	5.47	5.37	5.30	5.14	4.99
2.10	2.05	2.01	1.96	1.93	1.91	1.89	1.87	1.86	1.83	1.80
2.60	2.53	2.46	2.39	2.34	2.31	2.27	2.24	2.22	2.18	2.14
3.94	3.80	3.66	3.51	3.41	3.35	3.27	3.22	3.18	3.09	3.02
6.40	6.13	5.85	5.56	5.38	5.25	5.10	5.00	4.94	4.77	4.62
2.06	2.02	1.97	1.92	1.89	1.87	1.85	1.83	1.82	1.79	1.76
2.54	2.48	2.40	2.33	2.28	2.25	2.20	2.18	2.16	2.11	2.07
3.80	3.67	3.52	3.37	3.28	3.21	3.13	3.08	3.05	2.96	2.88
6.08	5.81	5.54	5.25	5.07	4.95	4.80	4.70	4.64	4.47	4.33
2.03	1.99	1.94	1.89	1.86	1.84	1.81	1.79	1.78	1.75	1.72
2.49	2.42	2.35	2.28	2.23	2.19	2.15	2.12	2.11	2.06	2.02
3.69	3.55	3.41	3.26	3.16	3.10	3.02	2.97	2.93	2.84	2.76
5.81	5.55	5.27	4.99	4.82	4.70	4.54	4.45	4.39	4.23	4.08
2.00	1.96	1.91	1.86	1.83	1.81	1.78	1.76	1.75	1.72	1.69
2.45	2.38	2.31	2.23	2.18	2.15	2.10	2.08	2.06	2.01	1.97
3.59	3.46	3.31	3.16	3.07	3.00	2.92	2.87	2.83	2.75	2.66
5.58	5.32	5.05	4.78	4.60	4.48	4.33	4.24	4.18	4.02	3.87
1.98	1.93	1.89	1.84	1.80	1.78	1.75	1.74	1.72	1.69	1.66
2.41	2.34	2.27	2.19	2.14	2.11	2.06	2.04	2.02	1.97	1.92
3.51	3.37	3.23	3.08	2.98	2.92	2.84	2.78	2.75	2.66	2.58
5.39	5.13	4.87	4.59	4.42	4.30	4.15	4.06	4.00	3.84	3.69
1.96	1.91	1.86	1.81	1.78	1.76	1.73	1.71	1.70	1.67	1.64
2.38	2.31	2.23	2.16	2.11	2.07	2.03	2.00	1.98	1.93	1.88
3.43	3.30	3.15	3.00	2.91	2.84	2.76	2.71	2.67	2.58	2.50
5.22	4.97	4.70	4.43	4.26	4.14	3.99	3.90	3.84	3.68	3.53
1.94	1.89	1.84	1.79	1.76	1.74	1.71	1.69	1.68	1.64	1.61
2.35	2.28	2.20	2.12	2.07	2.04	1.99	1.97	1.95	1.90	1.85
3.37	3.23	3.09	2.94	2.84	2.78	2.69	2.64	2.61	2.52	2.43
5.08	4.82	4.56	4.29	4.12	4.00	3.86	3.77	3.70	3.54	3.40
1.92	1.87	1.83	1.78	1.74	1.72	1.69	1.67	1.66	1.62	1.59
2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.94	1.92	1.87	1.82
3.31	3.17	3.03	2.88	2.79	2.72	2.64	2.58	2.55	2.46	2.37
4.95	4.70	4.44	4.17	4.00	3.88	3.74	3.64	3.58	3.42	3.28
1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.65	1.64	1.60	1.57
2.30	2.23	2.15	2.07	2.02	1.98	1.94	1.91	1.89	1.84	1.79
3.26	3.12	2.98	2.83	2.73	2.67	2.58	2.53	2.50	2.40	2.32
4.83	4.58	4.33	4.06	3.89	3.78	3.63	3.54	3.48	3.32	3.17
1.89	1.84	1.80	1.74	1.71	1.69	1.66	1.64	1.62	1.59	1.55
2.27	2.20	2.13	2.05	2.00	1.96	1.91	1.88	1.86	1.81	1.76
3.21	3.07	2.93	2.78	2.69	2.62	2.54	2.48	2.45	2.35	2.27
4.73	4.48	4.23	3.96	3.79	3.68	3.53	3.44	3.38	3.22	3.08
1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.62	1.61	1.57	1.54
2.25	2.18	2.11	2.03	1.97	1.94	1.89	1.86	1.84	1.79	1.74
3.17	3.03	2.89	2.74	2.64	2.58	2.49	2.44	2.40	2.31	2.22
4.64	4.39	4.14	3.87	3.71	3.59	3.45	3.36	3.29	3.14	2.99

(*continúa*)

Tabla A.9 Valores críticos de la distribución *F (continuación)*

ν_2 = denominador gl		ν_1 = numerador gl								
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
25	**.100**	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89
	.050	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
	.010	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
	.001	13.88	9.22	7.45	6.49	5.89	5.46	5.15	4.91	4.71
26	**.100**	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88
	.050	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
	.010	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
	.001	13.74	9.12	7.36	6.41	5.80	5.38	5.07	4.83	4.64
27	**.100**	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87
	.050	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
	.010	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
	.001	13.61	9.02	7.27	6.33	5.73	5.31	5.00	4.76	4.57
28	**.100**	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87
	.050	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
	.010	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
	.001	13.50	8.93	7.19	6.25	5.66	5.24	4.93	4.69	4.50
29	**.100**	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86
	.050	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
	.010	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
	.001	13.39	8.85	7.12	6.19	5.59	5.18	4.87	4.64	4.45
30	**.100**	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85
	.050	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
	.010	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
	.001	13.29	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.39
40	**.100**	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79
	.050	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
	.010	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
	.001	12.61	8.25	6.59	5.70	5.13	4.73	4.44	4.21	4.02
50	**.100**	2.81	2.41	2.20	2.06	1.97	1.90	1.84	1.80	1.76
	.050	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07
	.010	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78
	.001	12.22	7.96	6.34	5.46	4.90	4.51	4.22	4.00	3.82
60	**.100**	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74
	.050	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
	.010	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
	.001	11.97	7.77	6.17	5.31	4.76	4.37	4.09	3.86	3.69
100	**.100**	2.76	2.36	2.14	2.00	1.91	1.83	1.78	1.73	1.69
	.050	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97
	.010	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59
	.001	11.50	7.41	5.86	5.02	4.48	4.11	3.83	3.61	3.44
200	**.100**	2.73	2.33	2.11	1.97	1.88	1.80	1.75	1.70	1.66
	.050	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93
	.010	6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50
	.001	11.15	7.15	5.63	4.81	4.29	3.92	3.65	3.43	3.26
1000	**.100**	2.71	2.31	2.09	1.95	1.85	1.78	1.72	1.68	1.64
	.050	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.89
	.010	6.66	4.63	3.80	3.34	3.04	2.82	2.66	2.53	2.43
	.001	10.89	6.96	5.46	4.65	4.14	3.78	3.51	3.30	3.13

(*continúa*)

Tabla A.9 Valores críticos de la distribución *F (continuación)*

ν_1 = numerador gl										
10	**12**	**15**	**20**	**25**	**30**	**40**	**50**	**60**	**120**	**1000**
1.87	1.82	1.77	1.72	1.68	1.66	1.63	1.61	1.59	1.56	1.52
2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.84	1.82	1.77	1.72
3.13	2.99	2.85	2.70	2.60	2.54	2.45	2.40	2.36	2.27	2.18
4.56	4.31	4.06	3.79	3.63	3.52	3.37	3.28	3.22	3.06	2.91
1.86	1.81	1.76	1.71	1.67	1.65	1.61	1.59	1.58	1.54	1.51
2.22	2.15	2.07	1.99	1.94	1.90	1.85	1.82	1.80	1.75	1.70
3.09	2.96	2.81	2.66	2.57	2.50	2.42	2.36	2.33	2.23	2.14
4.48	4.24	3.99	3.72	3.56	3.44	3.30	3.21	3.15	2.99	2.84
1.85	1.80	1.75	1.70	1.66	1.64	1.60	1.58	1.57	1.53	1.50
2.20	2.13	2.06	1.97	1.92	1.88	1.84	1.81	1.79	1.73	1.68
3.06	2.93	2.78	2.63	2.54	2.47	2.38	2.33	2.29	2.20	2.11
4.41	4.17	3.92	3.66	3.49	3.38	3.23	3.14	3.08	2.92	2.78
1.84	1.79	1.74	1.69	1.65	1.63	1.59	1.57	1.56	1.52	1.48
2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.79	1.77	1.71	1.66
3.03	2.90	2.75	2.60	2.51	2.44	2.35	2.30	2.26	2.17	2.08
4.35	4.11	3.86	3.60	3.43	3.32	3.18	3.09	3.02	2.86	2.72
1.83	1.78	1.73	1.68	1.64	1.62	1.58	1.56	1.55	1.51	1.47
2.18	2.10	2.03	1.94	1.89	1.85	1.81	1.77	1.75	1.70	1.65
3.00	2.87	2.73	2.57	2.48	2.41	2.33	2.27	2.23	2.14	2.05
4.29	4.05	3.80	3.54	3.38	3.27	3.12	3.03	2.97	2.81	2.66
1.82	1.77	1.72	1.67	1.63	1.61	1.57	1.55	1.54	1.50	1.46
2.16	2.09	2.01	1.93	1.88	1.84	1.79	1.76	1.74	1.68	1.63
2.98	2.84	2.70	2.55	2.45	2.39	2.30	2.25	2.21	2.11	2.02
4.24	4.00	3.75	3.49	3.33	3.22	3.07	2.98	2.92	2.76	2.61
1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.48	1.47	1.42	1.38
2.08	2.00	1.92	1.84	1.78	1.74	1.69	1.66	1.64	1.58	1.52
2.80	2.66	2.52	2.37	2.27	2.20	2.11	2.06	2.02	1.92	1.82
3.87	3.64	3.40	3.14	2.98	2.87	2.73	2.64	2.57	2.41	2.25
1.73	1.68	1.63	1.57	1.53	1.50	1.46	1.44	1.42	1.38	1.33
2.03	1.95	1.87	1.78	1.73	1.69	1.63	1.60	1.58	1.51	1.45
2.70	2.56	2.42	2.27	2.17	2.10	2.01	1.95	1.91	1.80	1.70
3.67	3.44	3.20	2.95	2.79	2.68	2.53	2.44	2.38	2.21	2.05
1.71	1.66	1.60	1.54	1.50	1.48	1.44	1.41	1.40	1.35	1.30
1.99	1.92	1.84	1.75	1.69	1.65	1.59	1.56	1.53	1.47	1.40
2.63	2.50	2.35	2.20	2.10	2.03	1.94	1.88	1.84	1.73	1.62
3.54	3.32	3.08	2.83	2.67	2.55	2.41	2.32	2.25	2.08	1.92
1.66	1.61	1.56	1.49	1.45	1.42	1.38	1.35	1.34	1.28	1.22
1.93	1.85	1.77	1.68	1.62	1.57	1.52	1.48	1.45	1.38	1.30
2.50	2.37	2.22	2.07	1.97	1.89	1.80	1.74	1.69	1.57	1.45
3.30	3.07	2.84	2.59	2.43	2.32	2.17	2.08	2.01	1.83	1.64
1.63	1.58	1.52	1.46	1.41	1.38	1.34	1.31	1.29	1.23	1.16
1.88	1.80	1.72	1.62	1.56	1.52	1.46	1.41	1.39	1.30	1.21
2.41	2.27	2.13	1.97	1.87	1.79	1.69	1.63	1.58	1.45	1.30
3.12	2.90	2.67	2.42	2.26	2.15	2.00	1.90	1.83	1.64	1.43
1.61	1.55	1.49	1.43	1.38	1.35	1.30	1.27	1.25	1.18	1.08
1.84	1.76	1.68	1.58	1.52	1.47	1.41	1.36	1.33	1.24	1.11
2.34	2.20	2.06	1.90	1.79	1.72	1.61	1.54	1.50	1.35	1.16
2.99	2.77	2.54	2.30	2.14	2.02	1.87	1.77	1.69	1.49	1.22

Tabla A.10 Valores críticos para la distribución de rango estudentizado

		m										
		2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
5	.05	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.17	7.32
	.01	5.70	6.98	7.80	8.42	8.91	9.32	9.67	9.97	10.24	10.48	10.70
6	.05	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49	6.65	6.79
	.01	5.24	6.33	7.03	7.56	7.97	8.32	8.61	8.87	9.10	9.30	9.48
7	.05	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16	6.30	6.43
	.01	4.95	5.92	6.54	7.01	7.37	7.68	7.94	8.17	8.37	8.55	8.71
8	.05	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.05	6.18
	.01	4.75	5.64	6.20	6.62	6.96	7.24	7.47	7.68	7.86	8.03	8.18
9	.05	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74	5.87	5.98
	.01	4.60	5.43	5.96	6.35	6.66	6.91	7.13	7.33	7.49	7.65	7.78
10	.05	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60	5.72	5.83
	.01	4.48	5.27	5.77	6.14	6.43	6.67	6.87	7.05	7.21	7.36	7.49
11	.05	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49	5.61	5.71
	.01	4.39	5.15	5.62	5.97	6.25	6.48	6.67	6.84	6.99	7.13	7.25
12	.05	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39	5.51	5.61
	.01	4.32	5.05	5.50	5.84	6.10	6.32	6.51	6.67	6.81	6.94	7.06
13	.05	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32	5.43	5.53
	.01	4.26	4.96	5.40	5.73	5.98	6.19	6.37	6.53	6.67	6.79	6.90
14	.05	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.36	5.46
	.01	4.21	4.89	5.32	5.63	5.88	6.08	6.26	6.41	6.54	6.66	6.77
15	.05	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20	5.31	5.40
	.01	4.17	4.84	5.25	5.56	5.80	5.99	6.16	6.31	6.44	6.55	6.66
16	.05	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15	5.26	5.35
	.01	4.13	4.79	5.19	5.49	5.72	5.92	6.08	6.22	6.35	6.46	6.56
17	.05	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11	5.21	5.31
	.01	4.10	4.74	5.14	5.43	5.66	5.85	6.01	6.15	6.27	6.38	6.48
18	.05	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07	5.17	5.27
	.01	4.07	4.70	5.09	5.38	5.60	5.79	5.94	6.08	6.20	6.31	6.41
19	.05	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04	5.14	5.23
	.01	4.05	4.67	5.05	5.33	5.55	5.73	5.89	6.02	6.14	6.25	6.34
20	.05	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01	5.11	5.20
	.01	4.02	4.64	5.02	5.29	5.51	5.69	5.84	5.97	6.09	6.19	6.28
24	.05	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92	5.01	5.10
	.01	3.96	4.55	4.91	5.17	5.37	5.54	5.69	5.81	5.92	6.02	6.11
30	.05	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.72	4.82	4.92	5.00
	.01	3.89	4.45	4.80	5.05	5.24	5.40	5.54	5.65	5.76	5.85	5.93
40	.05	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.73	4.82	4.90
	.01	3.82	4.37	4.70	4.93	5.11	5.26	5.39	5.50	5.60	5.69	5.76
60	.05	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65	4.73	4.81
	.01	3.76	4.28	4.59	4.82	4.99	5.13	5.25	5.36	5.45	5.53	5.60
120	.05	2.80	3.36	3.68	3.92	4.10	4.24	4.36	4.47	4.56	4.64	4.71
	.01	3.70	4.20	4.50	4.71	4.87	5.01	5.12	5.21	5.30	5.37	5.44
∞	.05	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47	4.55	4.62
	.01	3.64	4.12	4.40	4.60	4.76	4.88	4.99	5.08	5.16	5.23	5.29

Tabla A.11 Áreas de cola de la curva chi-cuadrada

Área cola superior	$\nu = 1$	$\nu = 2$	$\nu = 3$	$\nu = 4$	$\nu = 5$
$> .100$	< 2.70	< 4.60	< 6.25	< 7.77	< 9.23
.100	2.70	4.60	6.25	7.77	9.23
.095	2.78	4.70	6.36	7.90	9.37
.090	2.87	4.81	6.49	8.04	9.52
.085	2.96	4.93	6.62	8.18	9.67
.080	3.06	5.05	6.75	8.33	9.83
.075	3.17	5.18	6.90	8.49	10.00
.070	3.28	5.31	7.06	8.66	10.19
.065	3.40	5.46	7.22	8.84	10.38
.060	3.53	5.62	7.40	9.04	10.59
.055	3.68	5.80	7.60	9.25	10.82
.050	3.84	5.99	7.81	9.48	11.07
.045	4.01	6.20	8.04	9.74	11.34
.040	4.21	6.43	8.31	10.02	11.64
.035	4.44	6.70	8.60	10.34	11.98
.030	4.70	7.01	8.94	10.71	12.37
.025	5.02	7.37	9.34	11.14	12.83
.020	5.41	7.82	9.83	11.66	13.38
.015	5.91	8.39	10.46	12.33	14.09
.010	6.63	9.21	11.34	13.27	15.08
.005	7.87	10.59	12.83	14.86	16.74
.001	10.82	13.81	16.26	18.46	20.51
$< .001$	> 10.82	> 13.81	> 16.26	> 18.46	> 20.51

Área cola superior	$\nu = 6$	$\nu = 7$	$\nu = 8$	$\nu = 9$	$\nu = 10$
$> .100$	< 10.64	< 12.01	< 13.36	< 14.68	< 15.98
.100	10.64	12.01	13.36	14.68	15.98
.095	10.79	12.17	13.52	14.85	16.16
.090	10.94	12.33	13.69	15.03	16.35
.085	11.11	12.50	13.87	15.22	16.54
.080	11.28	12.69	14.06	15.42	16.75
.075	11.46	12.88	14.26	15.63	16.97
.070	11.65	13.08	14.48	15.85	17.20
.065	11.86	13.30	14.71	16.09	17.44
.060	12.08	13.53	14.95	16.34	17.71
.055	12.33	13.79	15.22	16.62	17.99
.050	12.59	14.06	15.50	16.91	18.30
.045	12.87	14.36	15.82	17.24	18.64
.040	13.19	14.70	16.17	17.60	19.02
.035	13.55	15.07	16.56	18.01	19.44
.030	13.96	15.50	17.01	18.47	19.92
.025	14.44	16.01	17.53	19.02	20.48
.020	15.03	16.62	18.16	19.67	21.16
.015	15.77	17.39	18.97	20.51	22.02
.010	16.81	18.47	20.09	21.66	23.20
.005	18.54	20.27	21.95	23.58	25.18
.001	22.45	24.32	26.12	27.87	29.58
$< .001$	> 22.45	> 24.32	> 26.12	> 27.87	> 29.58

(*continúa*)

Tabla A.11 Áreas de cola de la curva chi-cuadrada *(continuación)*

Área cola superior	$\nu = 11$	$\nu = 12$	$\nu = 13$	$\nu = 14$	$\nu = 15$
$> .100$	< 17.27	< 18.54	< 19.81	< 21.06	< 22.30
.100	17.27	18.54	19.81	21.06	22.30
.095	17.45	18.74	20.00	21.26	22.51
.090	17.65	18.93	20.21	21.47	22.73
.085	17.85	19.14	20.42	21.69	22.95
.080	18.06	19.36	20.65	21.93	23.19
.075	18.29	19.60	20.89	22.17	23.45
.070	18.53	19.84	21.15	22.44	23.72
.065	18.78	20.11	21.42	22.71	24.00
.060	19.06	20.39	21.71	23.01	24.31
.055	19.35	20.69	22.02	23.33	24.63
.050	19.67	21.02	22.36	23.68	24.99
.045	20.02	21.38	22.73	24.06	25.38
.040	20.41	21.78	23.14	24.48	25.81
.035	20.84	22.23	23.60	24.95	26.29
.030	21.34	22.74	24.12	25.49	26.84
.025	21.92	23.33	24.73	26.11	27.48
.020	22.61	24.05	25.47	26.87	28.25
.015	23.50	24.96	26.40	27.82	29.23
.010	24.72	26.21	27.68	29.14	30.57
.005	26.75	28.29	29.81	31.31	32.80
.001	31.26	32.90	34.52	36.12	37.69
$< .001$	> 31.26	> 32.90	> 34.52	> 36.12	> 37.69

Área cola superior	$\nu = 16$	$\nu = 17$	$\nu = 18$	$\nu = 19$	$\nu = 20$
$> .100$	< 23.54	< 24.77	< 25.98	< 27.20	< 28.41
.100	23.54	24.76	25.98	27.20	28.41
.095	23.75	24.98	26.21	27.43	28.64
.090	23.97	25.21	26.44	27.66	28.88
.085	24.21	25.45	26.68	27.91	29.14
.080	24.45	25.70	26.94	28.18	29.40
.075	24.71	25.97	27.21	28.45	29.69
.070	24.99	26.25	27.50	28.75	29.99
.065	25.28	26.55	27.81	29.06	30.30
.060	25.59	26.87	28.13	29.39	30.64
.055	25.93	27.21	28.48	29.75	31.01
.050	26.29	27.58	28.86	30.14	31.41
.045	26.69	27.99	29.28	30.56	31.84
.040	27.13	28.44	29.74	31.03	32.32
.035	27.62	28.94	30.25	31.56	32.85
.030	28.19	29.52	30.84	32.15	33.46
.025	28.84	30.19	31.52	32.85	34.16
.020	29.63	30.99	32.34	33.68	35.01
.015	30.62	32.01	33.38	34.74	36.09
.010	32.00	33.40	34.80	36.19	37.56
.005	34.26	35.71	37.15	38.58	39.99
.001	39.25	40.78	42.31	43.81	45.31
$< .001$	> 39.25	> 40.78	> 42.31	> 43.81	> 45.31

Tabla A.12 Valores críticos para la prueba de normalidad Ryan-Joiner

		.10	.05	.01
n	**5**	.9033	.8804	.8320
	10	.9347	.9180	.8804
	15	.9506	.9383	.9110
	20	.9600	.9503	.9290
	25	.9662	.9582	.9408
	30	.9707	.9639	.9490
	40	.9767	.9715	.9597
	50	.9807	.9764	.9664
	60	.9835	.9799	.9710
	75	.9865	.9835	.9757

Tabla A.13 Valores críticos para la prueba Wilcoxon de rangos con signo

$P_0(S_+ \geq c_1) = P(S_+ \geq c_1$ cuando H_0 es verdadera)

n	c_1	$P_0(S_+ \geq c_1)$	n	c_1	$P_0(S_+ \geq c_1)$
3	6	.125		78	.011
4	9	.125		79	.009
	10	.062		81	.005
5	13	.094	14	73	.108
	14	.062		74	.097
	15	.031		79	.052
6	17	.109		84	.025
	19	.047		89	.010
	20	.031		92	.005
	21	.016	15	83	.104
7	22	.109		84	.094
	24	.055		89	.053
	26	.023		90	.047
	28	.008		95	.024
8	28	.098		100	.011
	30	.055		101	.009
	32	.027		104	.005
	34	.012	16	93	.106
	35	.008		94	.096
	36	.004		100	.052
9	34	.102		106	.025
	37	.049		112	.011
	39	.027		113	.009
	42	.010		116	.005
	44	.004	17	104	.103
10	41	.097		105	.095
	44	.053		112	.049
	47	.024		118	.025
	50	.010		125	.010
	52	.005		129	.005
11	48	.103	18	116	.098
	52	.051		124	.049
	55	.027		131	.024
	59	.009		138	.010
	61	.005		143	.005
12	56	.102	19	128	.098
	60	.055		136	.052
	61	.046		137	.048
	64	.026		144	.025
	68	.010		152	.010
	71	.005		157	.005
13	64	.108	20	140	.101
	65	.095		150	.049
	69	.055		158	.024
	70	.047		167	.010
	74	.024		172	.005

Tabla A.14 Valores críticos para la prueba Wilcoxon de suma de rangos

$P_0(W \geq c) = P(W \geq c$ cuando H_0 es verdadera)

m	n	c	$P_0(W \geq c)$	m	n	c	$P_0(W \geq c)$
3	3	15	.05			40	.004
	4	17	.057		6	40	.041
		18	.029			41	.026
	5	20	.036			43	.009
		21	.018			44	.004
	6	22	.048		7	43	.053
		23	.024			45	.024
		24	.012			47	.009
	7	24	.058			48	.005
		26	.017		8	47	.047
		27	.008			49	.023
	8	27	.042			51	.009
		28	.024			52	.005
		29	.012	6	6	50	.047
		30	.006			52	.021
4	4	24	.057			54	.008
		25	.029			55	.004
		26	.014		7	54	.051
	5	27	.056			56	.026
		28	.032			58	.011
		29	.016			60	.004
		30	.008		8	58	.054
	6	30	.057			61	.021
		32	.019			63	.01
		33	.010			65	.004
		34	.005	7	7	66	.049
	7	33	.055			68	.027
		35	.021			71	.009
		36	.012			72	.006
		37	.006		8	71	.047
	8	36	.055			73	.027
		38	.024			76	.01
		40	.008			78	.005
		41	.004	8	8	84	.052
5	5	36	.048			87	.025
		37	.028			90	.01
		39	.008			92	.005

Tabla A.15 Valores críticos para el intervalo Wilcoxon de rangos con signo

$(\bar{x}_{(n(n+1)/2-c+1)}, \bar{x}_{(c)})$

n	Nivel de confianza (%)	*c*	*n*	Nivel de confianza (%)	*c*	*n*	Nivel de confianza (%)	*c*
5	93.8	15	13	99.0	81	20	99.1	173
	87.5	14		95.2	74		95.2	158
6	96.9	21		90.6	70		90.3	150
	93.7	20	14	99.1	93	21	99.0	188
	90.6	19		95.1	84		95.0	172
7	98.4	28		89.6	79		89.7	163
	95.3	26	15	99.0	104	22	99.0	204
	89.1	24		95.2	95		95.0	187
8	99.2	36		90.5	90		90.2	178
	94.5	32	16	99.1	117	23	99.0	221
	89.1	30		94.9	106		95.2	203
9	99.2	44		89.5	100		90.2	193
	94.5	39	17	99.1	130	24	99.0	239
	90.2	37		94.9	118		95.1	219
10	99.0	52		90.2	112		89.9	208
	95.1	47	18	99.0	143	25	99.0	257
	89.5	44		95.2	131		95.2	236
11	99.0	61		90.1	124		89.9	224
	94.6	55	19	99.1	158			
	89.8	52		95.1	144			
12	99.1	71		90.4	137			
	94.8	64						
	90.8	61						

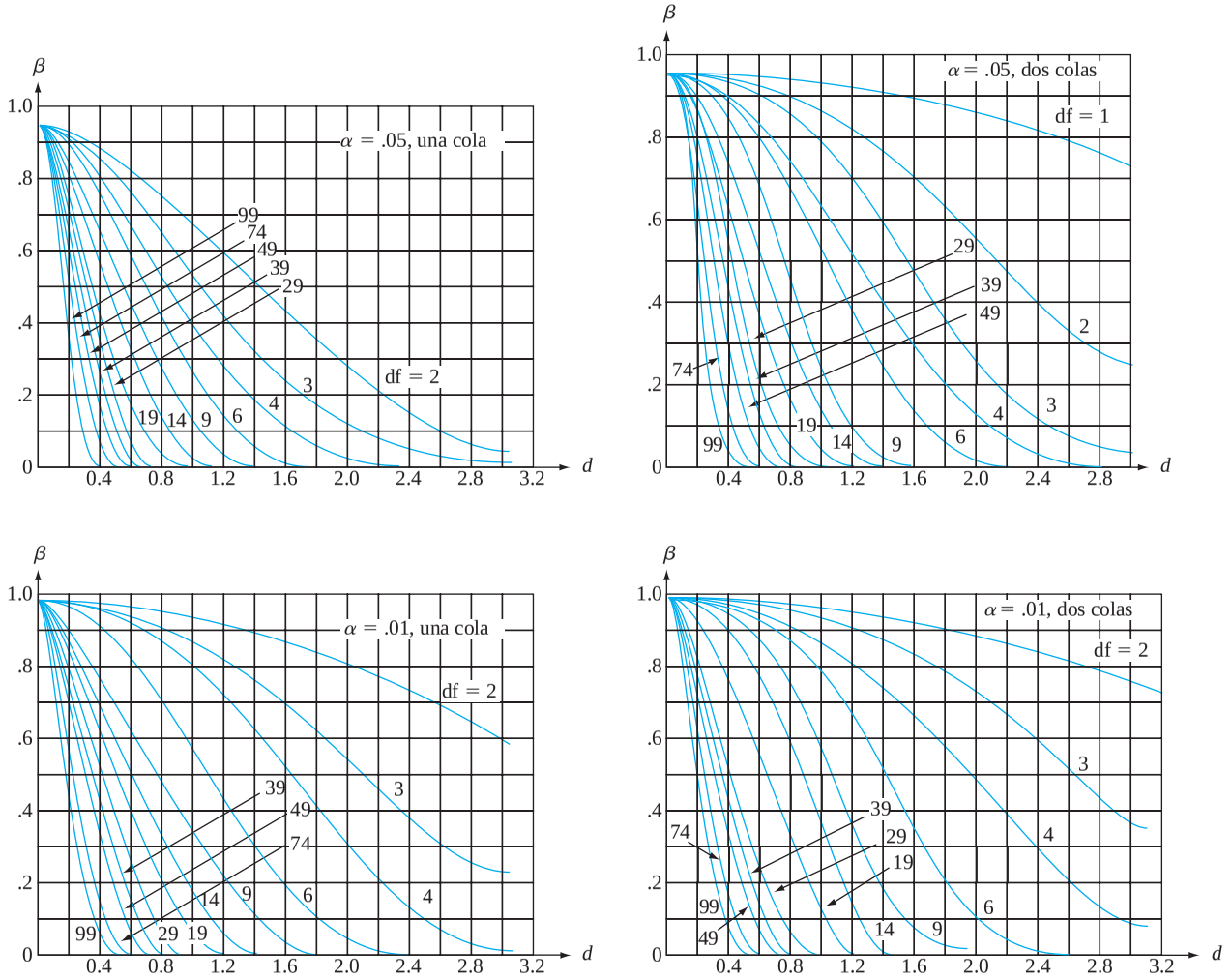
Tabla A.16 Valores críticos para el intervalo Wilcoxon de suma de rangos

$(d_{ij(mn-c+1)}, d_{ij(c)})$

Tamaño de muestra grande	Tamaño de muestra pequeña: 5		6		7		8	
	Nivel de confianza (%)	c	Nivel de confianza (%)	c	Nivel de confianza (%)	c	Nivel de confianza (%)	c
5	99.2	25						
	94.4	22						
	90.5	21						
6	99.1	29	99.1	34				
	94.8	26	95.9	31				
	91.8	25	90.7	29				
7	99.0	33	99.2	39	98.9	44		
	95.2	30	94.9	35	94.7	40		
	89.4	28	89.9	33	90.3	38		
8	98.9	37	99.2	44	99.1	50	99.0	56
	95.5	34	95.7	40	94.6	45	95.0	51
	90.7	32	89.2	37	90.6	43	89.5	48
9	98.8	41	99.2	49	99.2	56	98.9	62
	95.8	38	95.0	44	94.5	50	95.4	57
	88.8	35	91.2	42	90.9	48	90.7	54
10	99.2	46	98.9	53	99.0	61	99.1	69
	94.5	41	94.4	48	94.5	55	94.5	62
	90.1	39	90.7	46	89.1	52	89.9	59
11	99.1	50	99.0	58	98.9	66	99.1	75
	94.8	45	95.2	53	95.6	61	94.9	68
	91.0	43	90.2	50	89.6	57	90.9	65
12	99.1	54	99.0	63	99.0	72	99.0	81
	95.2	49	94.7	57	95.5	66	95.3	74
	89.6	46	89.8	54	90.0	62	90.2	70

Tamaño de muestra grande	Tamaño de muestra pequeña: 9		10		11		12	
	Nivel de confianza (%)	c	Nivel de confianza (%)	c	Nivel de confianza (%)	c	Nivel de confianza (%)	c
9	98.9	69						
	95.0	63						
	90.6	60						
10	99.0	76	99.1	84				
	94.7	69	94.8	76				
	90.5	66	89.5	72				
11	99.0	83	99.0	91	98.9	99		
	95.4	76	94.9	83	95.3	91		
	90.5	72	90.1	79	89.9	86		
12	99.1	90	99.1	99	99.1	108	99.0	116
	95.1	82	95.0	90	94.9	98	94.8	106
	90.5	78	90.7	86	89.6	93	89.9	101

Tabla A.17 Curvas β para pruebas t



Respuestas a ejercicios seleccionados de número impar

Capítulo 1

1. a. *Los Angeles Times, Oberlin Tribune, Gainesville Sun, Washington Post*
b. Duke Energy, Clorox, Seagate, Neiman Marcus
c. Vince Correa, Catherine Miller, Michael Cutler, Ken Lee
d. 2.97, 3.56, 2.20, 2.97

3. a. ¿Qué tan probable es que más de la mitad de las computadoras muestreadas necesiten o hayan necesitado servicio de garantía? ¿Cuál es el número esperado entre los 100 que necesitan servicio de garantía? ¿Qué tan probable es que el número que necesita servicio de garantía sea mayor que el número esperado en más de 10?
b. Suponga que 15 de las 100 muestreadas necesitaba servicio de garantía. ¿Qué confianza podemos tener de que la proporción de *todas* las computadoras que necesitan servicio de garantía sea entre .08 y .22? ¿La muestra proporciona evidencia obligatoria para concluir que más del 10% de esas computadoras necesitan servicio de garantía?

5. a. No. Todos los estudiantes que toman un curso grande de estadística que participan en un programa de IS de esta clase.
b. La aleatoriedad protege contra sesgos diversos, y ayuda a asegurar que los del grupo IS son tan semejantes como es posible a los estudiantes del grupo de control.
c. No habría base firme para evaluar la efectividad de la IS (nada a que las calificaciones de la IS pudieran compararse razonablemente).

7. Uno podría generar una muestra aleatoria sencilla de todas las casas unifamiliares de la ciudad, o una muestra aleatoria estratificada al tomar una muestra aleatoria sencilla de cada uno de los 10 vecindarios del distrito. De cada una de las casas seleccionadas, se determinarían valores de todas las variables deseadas. Éste sería un estudio enumerativo porque existe una población de objetos finita e identificable de la cual tomar muestras.

9. a. Posiblemente por error de medición, error de registro, diferencias en condiciones ambientales en el momento de la medición, etc.
b. No. No hay marco de muestreo.

11.

6L	430
6H	769689
7L	42014202
7H	
8L	011211410342
8H	9595578
9L	30
9H	58

La brecha en los datos, no hay calificaciones en los 70 altos.

13. a.
12 2 hoja: dígito de unidades
12 445
12 6667777
12 889999
13 00011111111
13 22222222233333333333333
13 44444444444444444555555555555555555
13 6666666666667777777777
13 888888888888999999
14 0000001111
14 2333333
14 444
14 77
simetría

b. Cerca de la forma de campana, centro ≈ 135, no hay dispersión insignificante, no hay brechas ni valores atípicos.

15.

Am		Fr
	8	1
157020153504	9	00645632
9324	10	2563
6306	11	6913
Tallo: centenas y decenas 058	12	325528
Hoja: unidades 8	13	7
	14	
	15	8
2	16	

Valores representativos: debajo de 100 para Am y debajo de 110 para Fr. Algo más de variabilidad en los tiempos de Fr que en horarios de Am. Sesgo positivo más extremo de Am que para Fr. 162 es un valor atípico Am y 158 es quizás un valor atípico para Fr.

17. a.

# No se apega	Frecuencia	Frec. rel.
0	7	.117
1	12	.200
2	13	.217
3	14	.233
4	6	.100
5	3	.050
6	3	.050
7	1	.017
8	1	.017
	60	1.001

b. .917, .867, 1 − .867 = .133
c. El histograma tiene un sesgo positivo considerable. Está centrado en un punto entre 2 y 3 y se dispersa bastante alrededor de su centro.

19. a. .99 (99%), .71 (71%) **b.** .64 (64%), .44 (44%)
c. Estrictamente hablando, el histograma no es unimodal, pero está cerca de serlo con un sesgo positivo moderado. Es probable que un tamaño muestral mucho más grande diera una imagen más uniforme.

21. a.

y	Frec.	Frec. rel.
0	17	.362
1	22	.468
2	6	.128
3	1	.021
4	0	.000
5	1	.021
	47	1.000

.362, .638

b.

z	Frec.	Frec. rel.
0	13	.277
1	11	.234
2	3	.064
3	7	.149
4	5	.106
5	3	.064
6	3	.064
7	0	.000
8	2	.043
	47	1.001

.894, .830

23. Los anchos de clase no son iguales, por lo que la escala de densidad debe ser utilizada. Las densidades de las seis clases son .2030, .1373, .0303, .0086, .0021 y .0009, respectivamente. El histograma resultante es unimodal con un sesgo positivo muy sustancial.

25.

Clase	Frec.	Clase	Frec.
10–<20	8	1.1–<1.2	2
20–<30	14	1.2–<1.3	6
30–<40	8	1.3–<1.4	7
40–<50	4	1.4–<1.5	9
50–<60	3	1.5–<1.6	6
60–<70	2	1.6–<1.7	4
70–<80	1	1.7–<1.8	5
	40	1.8–<1.9	1
			40

Original: sesgado positivamente;
Transformado: mucho más simétrico, no lejos de forma de campana.

27. a. La observación 50 cae en un límite de clase.
b.

Clase	Frec.	Frec. rel.
0–<50	9	.18
50–<100	19	.38
100–<150	11	.22
150–<200	4	.08
200–<300	4	.08
300–<400	2	.04
400–<500	0	.00
500–<600	1	.02
	50	1.00

Un valor representativo (central) está un poco abajo o un poco arriba de 100, dependiendo de cómo se mida el centro. Hay gran variabilidad en los tiempos de vida, en especial en valores del extremo superior de los datos. Hay varios candidatos para valores apartados.

c.

Clase	Frec.	Frec. rel.
2.25–<2.75	2	.04
2.75–<3.25	2	.04
3.25–<3.75	3	.06
3.75–<4.25	8	.16
4.25–<4.75	18	.36
4.75–<5.25	10	.20
5.25–<5.75	4	.08
5.75–<6.25	3	.06
	50	1.00

Hay mucha más simetría en la distribución de los valores $\ln(x)$ que en los valores x en sí, y menos variabilidad. Ya no hay brechas ni valores apartados obvios.
d. .38, .14

29.

Queja	Frec.	Frec. rel.
J	10	.1667
F	9	.1500
B	7	.1167
M	4	.0667
C	3	.0500
N	6	.1000
O	21	.3500
	60	1.0001

31.

Clase	Frec.	Frec. acum.	Frec. acumul. rel.
0–<4	2	2	.050
4–<8	14	16	.400
8–<12	11	27	.675
12–<16	8	35	.875
16–<20	4	39	.975
20–<24	0	39	.975
24–<28	1	40	1.000

33. **a.** 640.5, 582.5
b. 610.5, 582.5
c. 591.2
d. 593.71

35. **a.** $\bar{x} = 12.55$, $\tilde{x} = 12.50$, $\bar{x}_{tr(12.5)} = 12.40$. Borrar la observación más grande (18.0) hace que $\tilde{x}$ y $\bar{x}_{tr}$ sean un poco menores que $\bar{x}$.
b. En casi 4.0 **c.** No; multiplique los valores de $\bar{x}$ y $\tilde{x}$ por el factor de conversión 1/2.2.

37. $\bar{x}_{tr(10)} = 11.46$

39. **a.** $\bar{x} = 1.0297$, $\tilde{x} = 1.009$ **b.** .383

41. **a.** .7 **b.** También .7 **c.** 13

43. $\tilde{x} = 68.0$, $\bar{x}_{tr(20)} = 66.2$, $\bar{x}_{tr(30)} = 67.5$

45. **a.** $\bar{x} = 115.58$; las desviaciones son .82, .32, −.98, −.38, .22
b. .482, .694 **c.** .482 **d.** .482

47. $\bar{x} = 116.2$, $s = 25.75$. La magnitud de s indica una cantidad importante de variación alrededor del centro (una desviación "representativa" de casi 25).

49. **a.** 56.80, 197.8040 **b.** .5016, .708

51. **a.** 1264.766, 35.564 **b.** .351, .593

53. **a.** Bal: 1.121, 1.050, .536
Gr: 1.244, 1.100, .448
b. Proporciones típicas son muy similares para los dos tipos. Hay una variabilidad algo más en la muestra Bal, debido principalmente a los dos valores extremos (uno leve, un extremo). Para Bal, hay simetría sustancial en el centro 50%, pero la asimetría en general positivo. Para los de Gr, hay sesgo positivo sustancial en el centro 50% y un leve sesgo positivo en general.

55. **a.** 33 **b.** No
c. Un ligero sesgo positivo en la mitad central, pero más bien simétrico en su conjunto. La magnitud de variabilidad parece importante.
d. A lo sumo 32

57. **a.** Sí. 125.8 es un valor atípico extremo y 250.2 es un valor atípico moderado.
b. Además de la presencia de valores atípicos, hay sesgo positivo tanto en el 50% central de los datos como en general, excepto los valores atípicos. Con excepción de los dos valores atípicos, parece haber una cantidad relativamente pequeña de variabilidad en los datos.

59. **a.** DE (delirio excitado): .4, .10, 2.75, 2.65;
No-DE: 1.60, .30, 7.90, 7.60
b. DE: 8.9 y 9.2 son valores apartados moderados, y 11.7 y 21.0 son valores atípicos extremos.
No hay valores atípicos en la muestra sin delirio excitado.
c. Cuatro valores atípicos para DE, ninguno para no-DE. Sesgo positivo importante en ambas muestras; menos variabilidad en DE (para f_s menor), y las observaciones para no-DE tienden a ser un poco mayores que las observaciones con DE.

61. Valores atípicos, moderados y extremos, sólo a las 6 a.m. Las distribuciones a otras horas son bastante simétricas. La variabilidad aumenta un poco hasta las 2 p.m. y luego disminuye ligeramente, y lo mismo es cierto de valores "típicos" de coeficiente de vapor de gasolina.

63. $\bar{x} = 64.89$, $\tilde{x} = 64.70$, $s = 7.803$, baja $4^a = 57.8$, alta $4^a = 70.4$, $f_s = 12.6$. Un histograma que consta de 8 clases a partir de las 52, cada una de ancho de 4, es bimodal, pero cerca de unimodal con un sesgo positivo. Un diagrama de caja no muestra los valores extremos, hay un sesgo negativo muy leve en el centro 50% y el bigote superior es mucho más largo que el bigote más bajo.
b. .9231, .9053
c. .48

67. **a.** M:$\bar{x} = 3.64$, $\tilde{x} = 3.70$, $s = .269$, $f_s = .40$
F:$\bar{x} = 3.28$, $\tilde{x} = 3.15$, $s = .478$, $f_s = .50$
Los valores femeninos son por lo general un poco menores que los masculinos, y muestran un poco de más variabilidad. Una gráfica de caja M muestra sesgo negativo, mientras que una gráfica de caja F muestra sesgo positivo.
b. F:$\bar{x}_{tr(10)} = 3.24$ M:$\bar{x}_{tr(10)} = 3.652 \approx 3.65$

69. **a.** $\bar{y} = a\bar{x} + b$, $s_y^2 = a^2 s_x^2$ **b.** 189.14, 1.87

71. **a.** La media, mediana y media recortada son prácticamente idénticas, lo que sugiere una cantidad importante de simetría en la información; el hecho que los cuartiles estén a casi la misma distancia desde la mediana, y que las observaciones máximas sean casi equidistantes del centro, proporciona más apoyo para simetría. La desviación estándar es bastante pequeña con respecto a la media y mediana.
b. Vea los comentarios de (a). Además, usando 1.5(Q3 − Q1) como medida, las dos observaciones más grandes y las tres más pequeñas son valores atípicos moderados.

73. $\bar{x} = .9255$, $s = .0809$, $\tilde{x} = .93$, pequeña cantidad de variabilidad, un poco de sesgo

75. **a.** Los "resúmenes de cinco números" ($\tilde{x}$, los dos cuartos y las observaciones más pequeña y más grande) son idénticas y no hay valores apartados, de modo que las tres gráficas de caja individuales son idénticas.
b. Diferencias en variabilidad, naturaleza de brechas, y existencia de grupos para tres muestras.
c. No. Se pierde el detalle.

77. **c.** Las profundidades representativas son muy similares para los cuatro tipos de suelos, entre 1.5 y 2. Los datos de la C y suelos CL muestran mucho más variabilidad que los otros dos tipos. Los diagramas de caja para los tres primeros tipos muestran un importante sesgo positivo tanto en el centro 50% y en general. El diagrama de caja para el suelo SYCL muestra asimetría negativa en el centro 50% y un leve sesgo positivo en general. Por último, hay múltiples valores atípicos para los tres primeros tipos de suelos, incluyendo los valores atípicos extremos.

79. **a.** $\bar{x}_{n+1} = (n\bar{x}_n + x_{n+1})/(n + 1)$
c. 12.53, .532

81. Una asimetricidad positiva importante (suponiendo unimodalidad)

83. **a.** Todos los puntos caen en una recta de 45°. Los puntos caen debajo de la recta de 45°
b. Los puntos que caen debajo de la recta de 45°, indican una asimetricidad positiva importante.

Capítulo 2

1. **a.** $\mathcal{S}$ = {1324, 3124, 1342, 3142, 1423, 1432, 4123, 4132, 2314, 2341, 3214, 3241, 2413, 2431, 4213, 4231}
b. A = {1324, 1342, 1423, 1432}
c. B = {2314, 2341, 3214, 3241, 2413, 2431, 4213, 4231}
d. $A \cup B$ = {1324, 1342, 1423, 1432, 2314, 2341, 3214, 3241, 2413, 2431, 4213, 4231},
$A \cap B$ no contiene resultados (A y B son disjuntos),
A' = {3124, 3142, 4123, 4132, 2314, 2341, 3214, 3241, 2413, 2431, 4213, 4231}

3. **a.** A = {SSF, SFS, FSS}
b. B = {SSF, SFS, FSS, SSS}
c. C = {SFS, SSF, SSS}
d. C' = {FFF, FSF, FFS, FSS, SFF},
$A \cup C$ = {SSF, SFS, FSS, SSS},
$A \cap C$ = {SSF, SFS},
$B \cup C$ = {SSF, SFS, FSS, SSS} = B,
$B \cap C$ = {SSF, SFS, SSS} = C,

5. **a.** $\mathcal{S}$ = {(1, 1, 1), (1, 1, 2), (1, 1, 3), (1, 2, 1), (1, 2, 2), (1, 2, 3), (1, 3, 1), (1, 3, 2), (1, 3, 3), (2, 1, 1), (2, 1, 2), (2, 1, 3), (2, 2, 1), (2, 2, 2), (2, 2, 3), (2, 3, 1), (2, 3, 2), (2, 3, 3), (3, 1, 1), (3, 1, 2), (3, 1, 3), (3, 2, 1), (3, 2, 2), (3, 2, 3), (3, 3, 1), (3, 3, 2), (3, 3, 3)} **b.** {(1, 1, 1), (2, 2, 2), (3, 3, 3)} **c.** {(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), (3, 2, 1)} **d.** {(1, 1, 1), (1, 1, 3), (1, 3, 1), (1, 3, 3), (3, 1, 1), (3, 1, 3), (3, 3, 1), (3, 3, 3)}

7. **a.** Hay 35 resultados en $\mathcal{S}$. **b.** {*AABABAB, AABAABB, AAABBAB, AAABABB, AAAABBB*}

11. **a.** .07 **b.** .30 **c.** .57

13. **a.** .36 **b.** .64 **c.** .53
d. .47 **e.** .17 **f.** .75

15. **a.** .572 **b.** .879

17. **a.** Hay paquetes computarizados de estadística que no son SPSS y SAS.
b. .70 **c.** .80 **d.** .20

19. **a.** .8841 **b.** .0435

21. **a.** .10 **b.** .18, .19 **c.** .41 **d.** .59
e. .31 **f.** .69

23. **a.** .067 **b.** .400 **c.** .933 **d.** .533

25. **a.** .85 **b.** .15 **c.** .22 **d.** .35

27. **a.** .1 **b.** .7 **c.** .6

29. **a.** 676; 1296 **b.** 17,576; 46,656
c. 456,976; 1,679,616 **d.** .942

31. **a.** 243 **b.** 3645 días (casi 10 años)

33. **a.** 1,816,214,400 **b.** 659,067,881,572,000 **c.** 9,072,000

35. **a.** 38,760, .0048 **b.** .0054 **c.** .9946 **d.** .2885

37. **a.** 60 **b.** 10 **c.** .0456

39. **a.** .0839 **b.** .24975

41. **a.** 10,000 **b.** .9876 **c.** .0333 **d.** .0337

43. .000394, .00394, .00001539

45. **a.** .447, .500, .200 **b.** .400, .447 **c.** .211

47. **a.** .50 **b.** .50 **c.** .6125
d. .3875 **e.** .769

49. **a.** .34, .40 **b.** .588 **c.** .50

51. **a.** .436, **b.** .581

53. .083

55. .236

59. **a.** .21 **b.** .455 **c.** .264, .462, .274

61. **a.** .578, .278, .144 **b.** 0, .457, .543

63. **b.** .54 **c.** .68 **d.** .74 **e.** .7941

65. .087, .652, .261

67. .000329; muy inquietante.

69. **a.** .126 **b.** .05 **c.** .1125 **d.** .2725
e. .5325 **f.** .2113

71. **a.** .300 **b.** .820 **c.** .146

75. .401, .723

77. **a.** .00889 **b.** .00421

79. .0059

81. **a.** .95

83. **a.** .10, .20 **b.** 0

85. **a.** $p(2 - p)$ **b.** $1 - (1 - p)^n$ **c.** $(1 - p)^3$
d. $.9 + (1 - p)^3(.1)$
e. $.1(1 - p)^3/[.9 + .1(1 - p)^3] = .0137$ para $p = .5$

87. **a.** .40 **b.** .571
c. No: $.571 \neq .65$, y también $.40 \neq (.65)(.7)$ **d.** .733

89. $[2\pi(1 - \pi)]/(1 - \pi^2)$

91. **a.** .333, .444 **b.** .150 **c.** .291

93. .45, .32

95. **a.** .0083 **b.** .2 **c.** .2 **d.** .1074

97. .905

99. **a.** .974 **b.** .9754

101. .926

103. **a.** .018 **b.** .601

105. **a.** .883, .117 **b.** 23 **c.** .156

107. $1 - (1 - p_1)(1 - p_2) \cdot \cdots \cdot (1 - p_n)$

109. **a.** .0417 **b.** .375

111. $P(\text{contratar \#1}) = 6/24$ para $s = 0$, $= 11/24$ para $s = 1$, $= 10/24$ para $s = 2$, y $= 6/24$ para $s = 3$, de modo que $s = 1$ es mejor.

113. $1/4 = P(A_1 \cap A_2 \cap A_3)$
$\neq P(A_1) \cdot P(A_2) \cdot P(A_3) = 1/8$

Capítulo 3

1. $x = 0$ para *FFF*; $x = 1$ para *SFF*, *FSF*, y *FFS*; $x = 2$ para *SSF*, *SFS*, y *FSS*; y $x = 3$ para *SSS*

3. Z = promedio de los dos números, con posibles valores 2/2, 3/2, . . . , 12/2; W = valor absoluto de la diferencia, con valores posibles 0, 1, 2, 3, 4, 5

5. No. En el ejemplo 3.4, sea $Y = 1$ si a lo sumo se examinan 3 baterías y sea $Y = 0$ de otro modo. Entonces Y tiene sólo dos valores.

7. **a.** {0, 1, . . . , 12} ; discreta **c.** {1, 2, 3, . . .}; discreta
e. {0, c, 2c, . . . , 10,000c}, donde c es la regalía por libro; discreta **g.** $\{x: m \leq x \leq M\}$ donde $m(M)$ es la tensión mínima (máxima) posible; continua

9. **a.** {2, 4, 6, 8, . . .}, es decir, {2(1), 2(2), 2(3), 2(4), . . .}, una sucesión infinita; discreta
b. {2, 3, 4, 5, 6, . . .}, es decir, {1 + 1, 1 + 2, 1 + 3, 1 + 4, . . .}, una sucesión infinita, discreta

11. **a.** $p(4) = .45$, $p(6) = .40$, $p(8) = .15$, $p(x) = 0$ para $x \neq 4$, 6, u 8 **c.** .55, .15

13. **a.** .70 **b.** .45 **c.** .55
d. .71 **e.** .65 **f.** .45

15. **a.** (1, 2), (1, 3), (1, 4), (1, 5), (2, 3), (2, 4), (2, 5), (3, 4), (3, 5), (4, 5) **b.** $p(0) = .3$, $p(1) = .6$, $p(2) = .1$
c. $F(x) = 0$ para $x < 0$, $= .3$ para $0 \leq x < 1$, $= .9$ para $1 \leq x < 2$, e $= 1$ para $2 \leq x$

17. **a.** .81 **b.** .162 **c.** Es A; *AUUUA*, *UAUUA*, *UUAUA*, *UUUAA*; .0036

19. $p(0) = .09$, $p(1) = .40$, $p(2) = .32$, $p(3) = .19$

21. **b.** $p(x) = .301, .176, .125, .097, .079, .067, .058, .051, .046$ para $x = 1, 2, \ldots, 9$
c. $F(x) = 0$ para $x < 1$, $= .301$ para $1 \leq x < 2$, $= .477$ para $2 \leq x < 3, \ldots,$ $= .954$ para $8 \leq x < 9$, $= 1$ para $x \geq 9$
d. .602, .301

23. **a.** .20 **b.** .33 **c.** .78 **d.** .53

25. **a.** $p(y) = (1 - p)^y \cdot p$ para $y = 0, 1, 2, 3, \ldots$

27. **a.** 1234, 1243, 1324, . . . , 4321
b. $p(0) = 9/24$, $p(1) = 8/24$, $p(2) = 6/24$, $p(3) = 0$, $p(4) = 1/24$

29. **a.** 6.45 **b.** 15.6475 **c.** 3.96 **d.** 15.6475

31. 4.49, 2.12, .68

33. **a.** p **b.** $p(1 - p)$ **c.** p

35. $E[h_3(X)] = \$4.93$, $E[h_4(X)] = \$5.33$, así que 4 copias es mejor.

37. $E(X) = (n + 1)/2$, $E(X^2) = (n + 1)(2n + 1)/6$, $V(X) = (n^2 - 1)/12$

39. 6.1, .81, 88.5, 20.25

43. $E(X - c) = E(X) - c$, $E(X - \mu) = 0$

47. **a.** .515 **b.** .219 **c.** .012 **d.** .480
e. .965 **f.** .000 **g.** .595

49. **a.** .354 **b.** .114 **c.** .918

51. **a.** 6.25 **b.** 2.17 **c.** .030

53. **a.** .403 **b.** .787 **c.** .774

55. .1478

57. .407, independencia

59. **a.** .017 **b.** .811, .425 **c.** .006, .902, .586

61. Cuando $p = .9$, la probabilidad es .99 para A y .9963 para B. Si $p = .5$, estas probabilidades son .75 y .6875, respectivamente.

63. La tabulación para $p > .5$ es innecesaria.

65. **a.** 20, 16 **b.** 70, 21

67. $P(|X - \mu| \geq 2\sigma) = .042$ cuando $p = .5$ e $= .065$ cuando $p = .75$, en comparación con el límite superior de .25. Usando $k = 3$ en lugar de $k = 2$, estas probabilidades son .002 y .004, respectivamente, mientras que el límite superior es .11.

69. **a.** .114 **b.** .879 **c.** .121 **d.** Use la distribución binomial con $n = 15$, $p = .10$

71. **a.** $h(x; 15, 10, 20)$ para $x = 5, \ldots, 10$ **b.** .0325 **c.** .697

73. **a.** $h(x; 10, 10, 20)$ **b.** .033 **c.** $h(x; n, n, 2n)$

75. **a.** $nb(x; 2, .5)$ **b.** .188 **c.** .688 **d.** 2, 4

77. $nb(x; 6, .5)$, 6

79. **a.** .932 **b.** .065 **c.** .068 **d.** .492 **e.** .251

81. **a.** .011 **b.** .441 **c.** .554, .459 **d.** .945

83. Poisson(5) **a.** .492 **b.** .133

85. **a.** .122, .809, .283 **b.** 12, 3.464 **c.** .530, .011

87. **a.** .099 **b.** .135 **c.** 2

89. **a.** 4 **b.** .215 **c.** Al menos $-\ln(.1)/2 \approx 1.1513$ años

91. **a.** .221 **b.** 6,800,000 **c.** $p(x; 20.106)$

95. **b.** 3.114, .405, .636

97. **a.** $b(x; 15, .75)$ **b.** .686 **c.** .313 **d.** 11.25, 2.81 **e.** .310

99. .991

101. **a.** $p(x; 2.5)$ **b.** .067 **c.** .109

103. 1.813, 3.05

105. $p(2) = p^2$, $p(3) = (1 - p)p^2$, $p(4) = (1 - p)p^2$, $p(x) = [1 - p(2) - \cdots - p(x - 3)](1 - p)p^2$ para $x = 5, 6, 7, \ldots$; .99950841

107. **a.** .0029 **b.** .0767, .9702

109. **a.** .135 **b.** .00144 **c.** $\sum_{x=0}^{\infty}[p(x; 2)]^5$

111. 3.590

113. **a.** No **b.** .0273

115. **b.** $.5\mu_1 + .5\mu_2$ **c.** $.25(\mu_1 - \mu_2)^2 + .5(\mu_1 + \mu_2)$ **d.** .6 y .4 sustituya .5 y .5, respectivamente.

117. $\sum_{i=1}^{10}(p_{i+j+1} + p_{i-j-1})p_i$, donde $p_k = 0$ si $k < 0$ o $k > 10$.

121. **a.** 2.50 **b.** 3.1

Capítulo 4

1. **b.** .4625 **c.** .5, .2781

3. **b.** .5 **c.** .6875 **d.** .6328

5. **a.** .375 **b.** .125 **c.** .297 **d.** .578

7. **a.** $f(x) = .1$ para $25 \le x \le 35$ y 0 de otro modo **b.** .20 **c.** .40 **d.** .20

9. **a.** .562 **b.** .438, .438 **c.** .071

11. **a.** .25 **b.** .1875 **c.** .4375 **d.** 1.4142 **e.** $f(x) = x/2$ para $0 < x < 2$ **f.** 1.33 **g.** .222, .471 **h.** 2

13. **a.** 3 **b.** 0 para $x \le 1$, $1 - x^{-3}$ para $x > 1$ **c.** .125, .088 **d.** 1.5, .866 **e.** .924

15. **a.** $F(x) = 0$ para $x \le 0$, $= 90\left[\frac{x^9}{9} - \frac{x^{10}}{10}\right]$ para $0 < x < 1$, $= 1$ para $x \ge 1$ **b.** .0107 **c.** .0107, .0107 **d.** .9036 **e.** .818, .111 **f.** .3137

17. **a.** $A + (B - A)p$ **b.** $E(X) = (A + B)/2$, $\sigma_X = (B - A)/\sqrt{12}$ **c.** $[B^{n+1} - A^{n+1}]/[(n + 1)(B - A)]$

19. **a.** .597 **b.** .369 **c.** $f(x) = .3466 - .25\ln(x)$ para $0 < x < 4$

21. 314.79

23. 248, 3.60

25. **b.** 1.8(90avo percentil para X) + 32 **c.** a(Xpercentil) + b

27. 0, 1.814

29. **a.** 2.14 **b.** .81 **c.** 1.17 **d.** .97 **e.** 2.41

31. **a.** 2.54 **b.** 1.34 **c.** $-.42$

33. **a.** .9664 **b.** .2451 **c.** .8664

35. **a.** .3336 **b.** Aproximadamente 0 **c.** .5795 **d.** 6.524 **e.** .8028

37. **a.** 0, .5793, .5793 **b.** .3174, no **c.** < 87.6 o > 120.4

39. **a.** 36.7 **b.** 22.225 **c.** 3.179

41. .002

43. 10, .2

45. 7.3%

47. 21.155

49. **a.** .1190, .6969 **b.** .0021 **c.** .7054 **d.** > 5020 o < 1844 (usando $z_{.0005} = 3.295$) **e.** Normal, $\mu = 7.576$, $\sigma = 1.064$, .7054

51. .3174 para $k = 1$, .0456 para $k = 2$, .0026 para $k = 3$, en comparación con los límites de 1, .25, y .111, respectivamente.

53. **a.** Exacta: .212, .577, .573; Aproximada: .211, .567, .596
b. Exacta: .885, .575, .017; Aproximada: .885, .579, .012
c. Exacta: .002, .029, .617; Aproximada: .003, .033, .599

55. **a.** .9409 **b.** .9943

57. **b.** Normal, $\mu = 239$, $\sigma^2 = 12.96$

59. **a.** 1 **b.** 1 **c.** .982 **d.** .129

61. **a.** .480, .667, .147 **b.** .050, 0

63. **a.** corto $\Rightarrow$ plan #1 mejor, mientras que largo $\Rightarrow$ plan #2mejor
b. $1/\lambda = 10 \Rightarrow \mathrm{E}[h_1(X)] = 100$, $\mathrm{E}[h_2(X)] = 112.53$
$1/\lambda = 15 \Rightarrow \mathrm{E}[h_1(X)] = 150$, $\mathrm{E}[h_2(X)] = 138.51$

65. **a.** .238 **b.** .238 **c.** .313 **d.** .653 **e.** .653
f. .713

67. **a.** .424 **b.** .567, $\tilde{\mu} < 24$ **c.** 60 **d.** 66

69. **a.** $\cap A_i$ **b.** Exponencial con $\lambda = .05$
c. Exponencial con parámetro $n\lambda$

73. **a.** .826, .826, .0636 **b.** .664 **c.** 172.727

77. **a.** 123.97, 117.373 **b.** .5517 **c.** .1587

79. **a.** 9.164, .385 **b.** .8790 **c.** .4247, asimetricidad
d. No, ya que $P(X < 17{,}000) = .9332$

81. **a.** 149.157, 223.595 **b.** .9573 **c.** .0414
d. 148.41 **e.** 9.57 **f.** 125.90

83. $\alpha = \beta$

85. **b.** $[\Gamma(\alpha + \beta) \cdot \Gamma(m + \beta)]/[\Gamma(\alpha + \beta + m) \cdot \Gamma(\beta)]$, $\beta/(\alpha + \beta)$

87. Sí, porque el patrón en la gráfica es bastante lineal.

89. Sí

91. Sí

93. Graficar $\ln(x)$ vs. percentil z. El patrón es recto, de modo que es posible una distribución poblacional lognormal.

95. El patrón en la gráfica es bastante lineal; es muy posible que la resistencia sea distribuida normalmente.

97. Hay una curvatura importante en la gráfica. λ es un parámetro de escala (como es σ para la familia normal).

99. **a.** $F(y) = \frac{1}{48}(y^2 - y^3/18)$ para $0 \le y \le 12$
b. .259, .5, .241 **c.** 6, 43.2, 7.2
d. .518 **e.** 3.75

101. **a.** $f(x) = x^2$ para $0 \le x < 1$ e $= \frac{7}{4} - \frac{3}{4}x$ para $1 \le x \le \frac{7}{3}$ **b.** .917 **c.** 1.213

103. **a.** .9162 **b.** .9549 **c.** 1.3374

105. **a.** .3859 **b.** .1266 **c.** (72.97, 119.03)

107. **b.** $F(x) = 0$ para $x < -1$, $= (4x - x^3/3)/9 + \frac{11}{27}$ para $-1 \le x \le 2$, e $= 1$ para $x > 2$
c. No. $F(0) < .5 \Rightarrow \tilde{\mu} > 0$
d. $Y \sim \mathrm{Bin}(10, \frac{5}{27})$

109. **a.** .368, .828, .460 **b.** 352.53
c. $1/\beta \cdot \exp[-\exp(-(x - \alpha)/\beta)] \cdot \exp(-(x - \alpha)/\beta)$
d. α **e.** $\mu = 201.95$, moda $= 150$, $\tilde{\mu} = 182.99$

111. **a.** μ **b.** No **c.** 0 **d.** $(\alpha - 1)\beta$ **e.** $\nu - 2$

113. **b.** $p(1 - \exp(-\lambda_1 x)) + (1 - p)(1 - \exp(-\lambda_2 x))$ para $x \ge 0$ **c.** $p/\lambda_1 + (1 - p)/\lambda_2$
d. $V(X) = 2p/\lambda_1^2 + 2(1 - p)/\lambda_2^2 - \mu^2$
e. 1, $CV > 1$ **f.** $CV < 1$

115. **a.** Lognormal **b.** 1 **c.** 2.72, .0185

119. **a.** Exponencial con $\lambda = 1$
c. Gamma con parámetros α y $c\beta$

121. **a.** $(1/365)^3$ **b.** $(1/365)^2$ **c.** .000002145

123. **b.** Sea $u_1, u_2, u_3, \ldots$ una sucesión de observaciones de una distribución Unif[0, 1] (una sucesión de números aleatorios). Entonces con $x_i = (-.1)\ln(1 - u_i)$ las x_i son observaciones de una distribución exponencial con $\lambda = 10$.

125. $g(E(X)) \le E(g(X))$

127. **a.** 710, 84.423, .684 **b.** .376

Capítulo 5

1. **a.** .20 **b.** .42 **c.** Al menos una manguera está en uso en cada bomba; .70. **d.** $p_X(x) = .16, .34, .50$ para $x = 0, 1, 2$, respectivamente; $p_Y(y) = .24, .38, .38$ para $y = 0, 1, 2$, respectivamente; .50 **e.** No; $p(0, 0) \ne p_X(0) \cdot p_Y(0)$

3. **a.** .15 **b.** .40 **c.** .22 **d.** .17, .46

5. **a.** .054 **b.** .00018

7. **a.** .030 **b.** .120 **c.** .300 **d.** .380 **e.** Sí

9. **a.** 3/380,000 **b.** .3024 **c.** .3593
d. $10Kx^2 + .05$ para $20 \le x \le 30$ **e.** No

11. **a.** $e^{-\mu_1 - \mu_2} \cdot {\mu_1}^x \cdot {\mu_2}^y/x!y!$ **b.** $e^{-\mu_1 - \mu_2} \cdot [1 + \mu_1 + \mu_2]$
c. $e^{-(\mu_1 + \mu_2)} \cdot (\mu_1 + \mu_2)^m/m!$; Poisson $(\mu_1 + \mu_2)$

13. **a.** $e^{-x - y}$ para $x \ge 0, y \ge 0$ **b.** .400 **c.** .594
d. .330

15. **a.** $F(y) = 1 - e^{-\lambda y} + (1 - e^{-\lambda y})^2 - (1 - e^{-\lambda y})^3$ para $y \ge 0$
b. $2/3\lambda$

17. **a.** .25 **b.** .318 **c.** .637
d. $f_X(x) = 2\sqrt{R^2 - x^2}/\pi R^2$ para $-R \le x \le R$; no

19. **a.** $K(x^2 + y^2)/(10Kx^2 + .05)$; $K(x^2 + y^2)/(10Ky^2 + .05)$ **b.** .556, .549 **c.** 25.37, 2.87

21. **a.** $f(x_1, x_2, x_3)/f_{X_1, X_2}(x_1, x_2)$ **b.** $f(x_1, x_2, x_3)/f_{X_1}(x_1)$

23. .15

25. L^2

27. .25 h

29. $-\frac{2}{3}$

31. **a.** −.1082 **b.** −.0131

37. **a.**

$\bar{x}$	25	32.5	40	45	52.5	65
$p(\bar{x})$	.04	.20	.25	.12	.30	.09

$E(\bar{X}) = \mu = 44.5$

b.

s^2	0	112.5	312.5	800
$p(s^2)$	.38	.20	.30	.12

$E(S^2) = 212.25 = \sigma^2$

39.

Proporción	0	.1	.2	.3	.4	.5
Probabilidad	.000	.000	.000	.001	.005	.027

Proporción	.6	.7	.8	.9	1.0
Probabilidad	.088	.201	.302	.269	.107

41. **a.**

$\bar{x}$	1	1.5	2	2.5	3	3.5	4
$p(\bar{x})$	.16	.24	.25	.20	.10	.04	.01

b. .85 **c.**

r	0	1	2	3
$p(r)$	.30	.40	.22	.08

47. **a.** .6826 **b.** .1056

49. **a.** .6026 **b.** .2981

51. .7720

53. **a.** .0062 **b.** 0

55. **a.** .9838 **b.** .8926

57. .9616

59. **a.** .9986, .9986 **b.** .9875, .6266 **c.** .8357 **d.** .9525, .0003

61. **a.** 3.5, 2.27, 1.51 **b.** 15.4, 75.94, 8.71

63. **a.** .695 **b.** 4.0675 > 2.6775

65. **a.** .9232 **b.** .9660

67. .1588

69. **a.** 2400 **b.** 1205; independencia **c.** 2400, 41.77

71. **a.** 158, 430.25 **b.** .9788

73. **a.** Aproximadamente normal con media = 105, DE = 1.2649; aproximadamente normal con media = 100, DE = 1.0142 **b.** aproximadamente normal con media = 5, DE = 1.6213 **c.** .0068 **d.** .0010, sí

75. **a.** .2, .5, .3 para $x = 12, 15, 20$; .10, .35, .55 para $y = 12, 15, 20$ **b.** .25 **c.** No **d.** 33.35 **e.** 3.85

77. **a.** 3/81,250 **b.** $f_X(x) = k(250x - 10x^2)$ para $0 \le x \le 20$ e $= k(450x - 30x^2 + \frac{1}{2}x^3)$ para $20 < x \le 30$; $f_Y(y)$ resulta de sustituir y por x en $f_X(x)$. No son independientes. **c.** .355 **d.** 25.969 **e.** 204.6154, −.894 **f.** 7.66

79. ≈ 1

81. **a.** 400 min **b.** 70

83. 97

85. .9973

89. *b*, *c*. Chi cuadrada con $\nu = n$.

91. **a.** $\sigma_W^2/(\sigma_W^2 + \sigma_E^2)$ **b.** .9999

93. 26, 1.64

95. **a.** .6 **b.** $U = \rho X + \sqrt{1 - \rho^2}\, Y$

Capítulo 6

1. **a.** 8.14, $\overline{X}$ **b.** .77, $\tilde{X}$ **c.** 1.66, S **d.** .148 **e.** .204, $S/\overline{X}$

3. **a.** 1.348, $\overline{X}$ **b.** 1.348, $\overline{X}$ **c.** 1.781, $\overline{X} + 1.28S$ **d.** .6736 **e.** .0846

5. $N\bar{x} = 1{,}703{,}000$; $T - N\bar{d} = 1{,}591{,}300$; $T \cdot (\bar{x}/\bar{y}) = 1{,}601{,}438.281$

7. **a.** 120.6 **b.** 1,206,000 **c.** .80 **d.** 120.0

9. **a.** 2.11 **b.** .119

11. **b.** $\left[\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}\right]^{1/2}$ **c.** Use $\hat{p}_i = x_i/n_i$ y $\hat{q}_i = 1 - \hat{p}_i$ en lugar de p_i y q_i en el inciso (b) para $i = 1, 2$. **d.** −.245 **e.** .041

15. **a.** $\hat{\theta} = \sum X_i^2/2n$ **b.** 74.505

17. **b.** .444

19. **a.** $\hat{p} = 2\hat{\lambda} - .30 = .20$ **b.** $\hat{p} = (100\hat{\lambda} - 9)/70$

21. **b.** $\hat{\alpha} = 5$, $\hat{\beta} = 28.0/\Gamma(1.2)$

23. $\hat{\mu}_1 = \bar{x}$, $\hat{\mu}_2 = \bar{y}$, estimación de $(\mu_1 - \mu_2)$ es $\bar{x} - \bar{y}$.

25. **a.** 384.4, 18.86 **b.** 415.42

29. **a.** $\hat{\theta} =$ mín (X_i), $\hat{\lambda} = n/\sum[X_i -$ mín $(X_i)]$ **b.** .64, .202

33. Con x_i = tiempo entre nacimiento $i - 1$ y nacimiento i, $\hat{\lambda} = 6/\sum_{i=1}^{6} ix_i = .0436$.

35. 29.5

37. 1.0132

Capítulo 7

1. **a.** 99.5% **b.** 85% **c.** 2.96 **d.** 1.15

3. **a.** Más angosto **b.** No **c.** No **d.** No

5. **a.** (4.52, 5.18) **b.** (4.12, 5.00) **c.** .55 **d.** 94

7. Por un factor de 4; el ancho se reduce en un factor de 5.

9. **a.** $(\bar{x} - 1.645\sigma/\sqrt{n}, \infty)$; (4.57, ∞)
b. $(\bar{x} - z_\alpha \cdot \sigma/\sqrt{n}, \infty)$ **c.** $(-\infty, \bar{x} + z_\alpha \cdot \sigma/\sqrt{n})$; $(-\infty, 59.7)$

11. 950, .8714

13. **a.** (608.58, 699.74) **b.** 189

15. **a.** 80% **b.** 98% **c.** 75%

17. 134.53

19. (.513, .615)

21. $p < .273$ con 95% de confianza; sí

23. **a.** $p > .044$ con 95% de confianza
b. Si la misma fórmula se utiliza en la muestra tras muestra, en el largo plazo el valor real de p superará el 95% de los límites inferiores calculados.

25. **a.** 381 **b.** 339

29. **a.** 2.228 **b.** 2.086 **c.** 2.845 **d.** 2.680
e. 2.485 **f.** 2.571

31. **a.** 1.812 **b.** 1.753 **c.** 2.602 **d.** 3.747
e. 2.1716 (de Minitab) **f.** Alrededor de 2.43

33. **a.** Cantidad razonable de simetría, sin valores atípicos
b. Sí (con base en una gráfica normal de probabilidad)
c. (430.5, 446.1), sí, no

35. **a.** 95% de intervalo de confianza: (23.1, 26.9)
b. 95% IP: (17.2, 32.8), casi 4 veces el ancho

37. **a.** (.888, .964) **b.** (.752, 1.100) **c.** (.634, 1.218)

39. **a.** Sí **b.** (6.45, 98.01) **c.** (18.63, 85.83)

41. Todos 70%; (c) porque es más corto

43. **a.** 18.307 **b.** 3.940 **c.** .95 **d.** .10

45. (3.6, 8.1); no

47. **a.** 95% de intervalo de confianza: (6.702, 9.456)
b. (.166, .410)

49. **a.** Parece haber una ligera asimetría positiva en la mitad central de la muestra, pero el bigote inferior es mucho más largo que el bigote superior. La magnitud de la variabilidad es más bien importante, aun cuando no haya resultados atípicos.
b. Sí. La figura de puntos en una gráfica de probabilidad normal es razonablemente lineal.
c. (33.53, 43.79)

51. **a.** (.539, .581) **b.** 2398 **c.** No—97.5%

53. (−.84, −.16)

55. 246

57. $(2t_r/\chi^2_{1-\alpha/2,\,2r},\ 2t_r/\chi^2_{\alpha/2,\,2r}) = (65.3, 232.5)$

59. **a.** $(\text{máx}\,(x_i)/(1-\alpha/2)^{1/n},\ \text{máx}\,(x_i)/(\alpha/2)^{1/n})$
b. $(\text{máx}\,(x_i), \text{máx}\,(x_i)/\alpha^{1/n})$ **c.** (b); (4.2, 7.65)

61. (73.6, 78.8) contra (75.1, 79.6)

Capítulo 8

1. **a.** Sí **b.** No **c.** No
d. Sí **e.** No **f.** Sí

5. H_0: $\sigma = 0.5$ contra H_a: $\sigma < 0.5$. I: concluya que la variabilidad en grosor es satisfactoria cuando no lo es. II: concluya que la variabilidad en grosor no es satisfactoria cuando de hecho lo es.

7. I: concluyendo que la planta no se apega cuando sí se cumple; II: concluyendo que la planta se apega cuando de hecho no se cumple.

9. **a.** R_1 **b.** I: juzgando que una de las dos compañías es favorecida sobre la otra cuando ese no es el caso; II: juzgando que ninguna compañía es favorecida sobre la otra cuando de hecho las dos son realmente preferidas. **c.** .044
d. $\beta(.3) = \beta(.7) = .488$, $\beta(.4) = \beta(.6) = .845$
e. Rechazar H_0 a favor de H_a.

11. **a.** H_0: $\mu = 10$ contra H_a: $\mu \neq 10$ **b.** .01
c. .5319, .0078 **d.** 2.58
e. 10.1032 se sustituye con 10.124, y 9.8968 se sustituye con 9.876. **f.** $\bar{x} = 10.020$, de modo que H_0 no debe ser rechazada.
g. $z \geq 2.58$ o ≤ -2.58

13. **b.** .0004, 0, menor que .01

15. **a.** .0301 **b.** .003 **c.** .004

17. **a.** $z = 2.56 \geq 2.33$, y rechazar H_0. **b.** .8413 **c.** 143
d. .0052

19. **a.** $z = -2.27$, y no rechazar H_0. **b.** .2266 **c.** 22

21. **a.** $t_{.025,12} = 2.179 > 1.6$, y no rechazar H_0: $\mu = .5$.
b. $-1.6 > -2.179$, y no rechazar H_0.
c. No rechazar H_0.
d. Rechazar H_0 a favor de H_a: $\mu \neq .5$.

23. $t = 2.24 \geq 1.708$ y H_0 debe rechazarse. La información no sugiere una contradicción de creencia anterior.

25. **a.** $z = -3.33 \leq -2.58$ y rechazar H_0.
b. .1056 **c.** 217

27. **a.** $\bar{x} = .750$, $\tilde{x} = .640$, $s = .3025$, $f_s = .480$. Una gráfica de caja muestra importante asimetría positiva; no hay valores atípicos.
b. No. Una gráfica de probabilidad normal muestra curvatura importante. No, porque n es grande.
c. $z = -5.79$; rechazar H_0 a cualquier nivel razonable de significación; sí. **d.** .821

29. **a.** No, ya que $1.19 < 1.796$ **b.** .30 (con el software)

31. No, ya que $1.04 < 2.132$

33. **a.** No, ya que $2.44 < 2.539$ **b.** Sí, tipo II
c. .66 (con el software)

37. **a.** No, ya que $1.28 < 1.645$
b. I: dice que más de 20% son obesos cuando este no es el caso; II: concluye que 20% son obesos cuando el porcentaje real excede 20%.
c. .121

39. $z = 3.67 \geq 2.58$, y rechazar H_0: $p = .40$. No.

41. **a.** $z = -1.0$, así que no hay suficiente evidencia para concluir que $p < .25$; Por tanto se usan tapones de rosca.
b. I: No usar tapones de rosca cuando se justifica su uso;
II: Usar tapones de rosca cuando no se justifica su uso.

43. **a.** $z = 3.07 \geq 2.58$, rechazar H_0 y la premisa de la compañía.
b. .0332

45. No, no, sí. $R = \{5, 6, \ldots, 24, 25\}$, $\alpha = .098$, $\beta = .090$

47. **a.** Rechazar H_0. **b.** Rechazar H_0. **c.** No rechazar H_0.
d. Rechazar H_0. (una llamada cercana) **e.** No rechazar H_0.

49. **a.** .0778 **b.** .1841 **c.** .0250
d. .0066 **e.** .5438

51. **a.** .040 **b.** .018 **c.** .130 **d.** .653
e. $< .005$ **f.** $\approx .000$

53. Valor $P > \alpha$, y no rechazar H_0; no hay diferencia aparente.

55. Valor $P < .0004 < .01$, por lo que H_0: $\mu = 5$ podría ser rechazada a favor de H_a: $\mu \neq 5$.

57. No, ya que el valor $P = .2266$

59. **a.** Sí **b.** El valor P excede ligeramente .10, así que H_0: $\mu = 100$ no debería rechazarse y podría usarse el concreto.

61. $t \approx 1.9$, de modo que el valor $P \approx .116$. Por lo tanto, H_0 no debe ser rechazada.

63. **a.** .8980, .1049, .0014 **b.** Valor $P \approx 0$. Sí. **c.** No

65. $z = -3.12 \leq -1.96$, de modo que H_0 debe ser rechazada.

67. **a.** H_0: $\mu = .85$ contra H_a: $\mu \neq .85$
b. H_0 no puede ser rechazada por ninguna α.

69. **a.** No, porque el valor $P = .02 > .01$; sí, porque 45.31 excede por mucho 20, pero n es muy pequeña.
b. $\beta = .3$ (con software)

71. **a.** No; no
b. No, porque $z = .44$ y valor $P = .33 > .10$.

73. **a.** Aproximadamente .6; aproximadamente .2 (de la tabla A.17 del apéndice) **b.** $n = 28$

75. **a.** $z = 1.64 < 1.96$, so H_0 de modo que H_0 no puede ser rechazada; tipo II
b. .10. Sí.

77. Sí. $z = -3.32 \leq -3.08$, de modo que H_0 debe rechazarse.

79. No, porque $z = 1.33 < 2.05$.

81. Sí, porque $z = 4.4$ y el valor $P = 0 < .05$

83. **a.** $.01 <$ valor $P < .025$, de modo que no se rechace H_0; no hay contradicción

85. **a.** Para H_a: $\mu < \mu_0$, rechazar H_0 si $2\sum x_i/\mu_0 \leq \chi^2_{1-\alpha,\,2n}$
b. Valor estadístico de prueba $= 19.65 > 8.260$, y no rechazar H_0.

87. **a.** Sí, $\alpha = .002$

Capítulo 9

1. **a.** $-.4$ h; no depende **b.** .0724, .2691 **c.** No

3. $z = 1.76 < 2.33$, de modo que no rechace H_0.

5. **a.** $z = -2.90$, y rechace H_0. **b.** .0019
c. .8212 **d.** 66

7. No, ya que el valor de P para la prueba de 2 colas de .0602.

9. **a.** 6.2; sí **b.** $z = 1.14$, valor $P \approx .25$, no
c. No **d.** Un intervalo de confianza de 95% es (10.0, 21.8).

11. Un intervalo de confianza de 95% es (.99, 2.41).

13. 50

15. **b.** Aumenta.

17. **a.** 17 **b.** 21 **c.** 18 **d.** 26

19. $t = -1.20 > -t_{.01,9} = -2.764$, no rechazar H_0.

21. Puesto que no es verdad que; $-2.46 \leq -2.602$, no rechazamos H_0. No se tiene suficiente evidencia.

23. **b.** No **c.** $t = -.38 > -t_{\alpha/2,\,10}$ para cualquier α razonable, entonces no rechazar H_0 (valor $P \approx .7$).

25. (.3, 6.1), sí, sí

27. **a.** Intervalo de confianza de 99%: (.33, .71) **b.** Intervalo de confianza de 99%: $(-.07, .41)$, así que 0 es un posible valor de la diferencia.

29. $t = -2.10$, gl $= 25$, valor $P = .023$ d. Al nivel de significancia .05, concluiríamos que la cola resulta en un promedio de resistencia más alto, pero no al nivel de significancia de .01.

31. **a.** Centros prácticamente idénticos, sustancialmente más variabilidad en observaciones de rango medio que en observaciones de rango más alto
b. $(-7.9, 9.6)$, con base en 23 grados de libertad; no

33. $t = 1.33$, valor $P = .094$, no rechazar H_0, no

35. $t = -2.2$, grados de libertad $= 16$, valor $P = .021 > .01 = \alpha$, de modo que no se rechace H_0.

37. **a.** $(-.561, -.287)$ **b.** Entre -1.224 y $.376$

39. **a.** Sí. **b.** $t = 2.7$, valor $P = .018 < .05 = \alpha$, de modo que se rechace H_0.

41. **a.** $(-3.85, 11.35)$ **b.** Sí, porque el valor $P = .02$, al nivel de .05 puede parecer que se incrementa, pero no al nivel de .01. **c.** (7.02, 10.06)

43. **a.** No **b.** -49.1 **c.** 49.1

45. **a.** Sí, debido al patrón lineal en una gráfica de probabilidad normal. **b.** No, los datos están pareados, no son muestras independientes. **c.** $t = 3.66$, valor $P = .001$ (no .003), misma conclusión.

47. **a.** Un intervalo de confianza de 95%: $(-2.52, 1.05)$; posible que sean idénticos
b. El patrón lineal en prueba no pasada implica que la distribución de diferencia en normalidad es posible.

49. H_0 es rechazada porque $-4.18 \leq -2.33$

51. Valor $P = .4247$, de modo que H_0 no puede ser rechazada.

53. **a.** $z = .80 < 1.96$, entonces no rechazar H_0. **b.** $n = 1211$

55. **a.** El intervalo de confianza para $\ln(\theta)$ es $\ln(\hat{\theta}) \pm z_{\alpha/2}[(m - x)/(mx) + (n - y)/(ny)]^{1/2}$. Al tomar los antilogaritmos de los límites inferior y superior da un intervalo de confianza para θ misma.
b. (1.43, 2.31); parece que la aspirina es benéfica.

57. $(-.35, .07)$

59. **a.** 3.69 **b.** 4.82 **c.** .207 **d.** .271
e. 4.30 **f.** .212 **g.** .95 **h.** .94

61. $f = .384$; como $.167 < .384 < 3.63$, no rechazar H_0.

63. $f = 2.85 \geq 2.08$, así que rechace H_0; no parece haber más variabilidad en aumento de peso en dosis baja.

65. $(s_2^2F_{1-\alpha/2}/s_1^2, s_2^2F_{\alpha/2}/s_1^2)$; (.023, 1.99)

67. No. $t = 3.2$, gl $= 15$, valor $P = .006$, y rechace H_0: $\mu_1 - \mu_2 = 0$ usando ya sea $\alpha = .05$ o .01.

69. $z > 0 \Rightarrow$ valor $P > .5$, así que H_0: $p_1 - p_2 = 0$ no puede ser rechazada.

71. $(-299.3, 1517.9)$

73. Parecen diferir, ya que gl $= 14$, $t = -5.19$, valor $P = 0$.

75. Sí, $t = -2.25$, gl $= 57$, valor $P \approx .028$.

77. **a.** No. $t = -2.84$, gl $= 18$, valor $P \approx .012$
b. No. $t = -.56$, valor $P \approx .29$

79. $t = 3.9$, valor $P = .004$, así H_0 es rechazada al nivel de significación .05 o .01.

81. No, ni debe usarse la prueba t de dos muestras, porque una gráfica de probabilidad normal sugiere que la distribución de buena visibilidad no es normal.

83. No agrupado: gl $= 15$, $t = -1.8$, valor $P \approx .092$
Agrupado: gl $= 24$, $t = -1.9$, valor $P \approx .070$

85. **a.** $m = 141$, $n = 47$ **b.** $m = 240$, $n = 160$

87. No, $z = .83$, valor $P \approx .20$

89. .9015, .8264, .0294, .0000; promedios verdaderos de los cuartiles; no

91. Sí; $z = 4.2$, valor $P \approx 0$

93. **a.** Sí. $t = -6.4$, gl $= 57$, y valor $P \approx 0$
b. $t = 1.1$, valor $P = .14$, y no rechazar H_0.

95. $(-1.29, -.59)$

Capítulo 10

1. **a.** $f = 1.85 < 3.06 = F_{.05,\,4,\,15}$, de modo que no rechace H_0.
b. Valor $P > .10$

3. $f = 1.30 < 2.57 = F_{.10,\,2,\,21}$, de modo que el valor $P > .10$. H_0 no puede ser rechazada a ningún nivel razonable de significación.

5. $f = 1.73 < 5.49 = F_{.01,\,2,\,27}$, por lo que no parecen diferir los tres grados.

7. $f = 51.3$, P-value $= 0$, por lo que el valor $P > .10$. H_0 no puede ser rechazada a ningún nivel razonable de significación.

9. $f = 3.96$ y $F_{.05,\,3,\,20} = 3.10 < 3.96 < 4.94 = F_{.01,3,20}$, así que $.01 <$ valor $P < .05$. Entonces H_0 puede ser rechazada al nivel de significación .05; parece haber diferencias entre los granos.

11. $w = 36.09$

3	1	4	2	5
437.5	462.0	469.3	512.8	532.1

Las marcas 2 y 5 parecen no diferir, ni parece haber diferencia entre las marcas 1,3 y 4 pero cada marca del primer grupo parece diferir de manera importante respecto de todas las marcas del segundo grupo.

13.

3	1	4	2	5
427.5	462.0	469.3	502.8	532.1

15.

14.18	17.94	18.00	18.00	25.74	27.67

17. $(-.029, .379)$

19. Funcionará cualquier valor de la suma de los cuadrados del error (SSE) entre 422.16 y 431.88.

21. **a.** $f = 22.6$ y $F_{.01,\,5,\,78} \approx 3.3$, de modo que rechace H_0.
b. $(-99.16, -35.64)$, $(29.34, 94.16)$

23.

	1	2	3	4
1	–	2.88 ± 5.81	7.43 ± 5.81	12.78 ± 5.48
2	–	–	4.55 ± 6.13	9.90 ± 5.81
3	–	–	–	5.35 ± 5.81
4	–	–	–	–

4	3	2	1

25. a. Normal, varianzas iguales
b. SSTr = 8.33, SSE = 77.79, $f = 1.7$, H_0 no debe ser rechazada (valor $P > .10$)

27. a. $f = 3.75 \geq 3.10 = F_{.05,\,3,\,20}$, así que las marcas parecen diferir.

b. La normalidad es bastante posible (una gráfica de probabilidad normal de los residuos $x_{ij} - \bar{x}_{i\cdot}$ muestra un patrón lineal).

c. 4 3 2 1 Sólo las marcas 1 y 4 parecen diferir de manera importante.

31. Aproximadamente .62

33. arcsen $(\sqrt{x/n})$

35. a. $3.68 < 4.94$, así que H_0 no es rechazada.

b. $.029 > .01$, y otra vez H_0 no es rechazada.

37. $f = 8.44 > 6.49 = F_{.001}$, así el valor $P < .001$ y H_0 debe ser rechazada.

5 3 1 4 2 Este patrón de subrayar es un poco difícil de interpretar.

39. El intervalo de confianza es $(-.144, .474)$, que no incluye 0.

41. $f = 3.96 < 4.07$, así que H_0: $\sigma_A^2 = 0$ no puede ser rechazada.

43. $(-3.70, 1.04)$, $(-4.83, -.33)$, $(-3.77, 1.27)$, $(-3.99, .15)$. Sólo $\mu_1 - \mu_3$ entre estos cuatro contrastes parece diferir de modo importante de cero.

45. Son idénticos.

Capítulo 11

1. a. $f_A = 1.55$, así no se rechaza H_{0A}.
b. $f_B = 2.98$, así no se rechaza H_{0B}.

3. a. $f_A = 105.3 \geq F_{.01,3.9}$, así se concluye que hay un efecto de gasto de gas; $f_B = 13.0$, así se concluye que hay un efecto de gasto de líquido.

b. $w = 95.44$; 231.75 325.25 441.0 613.25, y sólo los dos gastos más bajos no difieren de modo importante entre sí.

c. 336.75 382.25 419.25 473 y sólo los gastos más bajo y más alto parecen diferir de modo importante entre sí.

5. $f_A = 2.56$, $F_{.01,3,12} = 5.95$, y parece no haber efecto debido al ángulo de tirón.

7. a.

Causa	gl	SC	CM	f
Tratamientos	2	28.78	14.39	1.04
Bloques	17	2977.67	175.16	12.68
Error	34	469.55	13.81	
Total	53	3476.00		

El promedio verdadero de adaptación no parece depender de qué tratamiento se administra. **b.** Sí, f_B es bastante grande, lo que sugiere una gran variabilidad entre los sujetos.

9.

Causa	gl	SC	CM	f	$F_{.05}$
Tratamientos	3	81.19	27.06	22.4	3.01
Bloques	8	66.50	8.31		
Error	24	29.06	1.21		
Total	35	176.75			

1	4	3	2
8.56	9.22	10.78	12.44

11. El patrón de la gráfica de probabilidad normal es bastante lineal. No hay un patrón discernible en una gráfica de residuos contra valores ajustados.

13. b. Cada SC se multiplica por c^2, pero f_A y f_B no cambian.

15. a. Aproximadamente .20, .43 **b.** Aproximadamente .30

17. a. $f_A = 3.76$, $f_B = 6.82$, $f_{AB} = .74$, y $F_{.05,2,9} = 4.26$, así que la cantidad de adición de fibra de carbono parece importante.

b. $f_A = 6.54$, $f_B = 5.33$, $f_{AB} = .27$

19. a.

Causa	gl	SC	CM	f
Carbón	2	1.00241	.50121	29.49
NaOH	2	.12431	.06216	3.66
Interacción	4	.01456	.00364	.21
Error	9	.15295	.01699	
Total	17	1.29423		

El tipo de carbón parece afectar la acidez total.

b. Los carbones 1 y 3 no difieren entre sí de modo importante, pero ambos difieren significativamente del carbón 2.

21. a, b.

Causa	gl	SC	CM	f
A	2	22,941.80	11,470.90	22.98
B	4	22,765.53	5691.38	5.60
AB	8	3993.87	499.23	.49
Error	15	15,253.50	1016.90	
Total	29	64,954.70		

H_{0A} y H_{0B} son rechazadas.

23.

Causa	gl	SC	CM	f
A	2	11,573.38	5786.69	$\frac{\text{MSA}}{\text{MSAB}} = 26.70$
B	4	17,930.09	4482.52	$\frac{\text{MSB}}{\text{MSE}} = 28.51$
AB	8	1734.17	216.77	$\frac{\text{MSAB}}{\text{MSE}} = 1.38$
Error	30	4716.67	157.22	
Total	44	35,954.31		

Como $F_{.01,8,30} = 3.17$, $F_{.01,2,8} = 8.65$, y $F_{.01,4,30} = 4.02$, H_{0G} no es rechazada pero H_{0A} y H_{0B} son rechazados.

25. $(-.373, -.033)$

27. a.

Causa	gl	SC	CM	f	$F_{.05}$
A	2	14,144.44	7072.22	61.06	3.35
B	2	5511.27	2755.64	23.79	3.35
C	2	244,696.39	122,348.20	1056.27	3.35

AB	4	1069.62	267.41	2.31	2.73
AC	4	62.67	15.67	.14	2.73
BC	4	331.67	82.92	.72	2.73
ABC	8	1080.77	135.10	1.17	2.31
Error	27	3127.50	115.83		
Total	53	270,024.33			

d. $Q_{.05,3,27} = 3.51, w = 8.90$, y los tres niveles difieren entre sí de manera importante.

29. a.

Causa	gl	SC	CM	*f*
A	2	12.896	6.448	1.04
B	1	100.041	100.041	16.10
C	3	393.416	131.139	21.10
AB	2	1.646	.823	<1
AC	6	71.021	11.837	1.905
BC	3	1.542	.514	<1
ABC	6	9.771	1.629	<1
Error	72	447.500	6.215	
Total	95	1037.833		

b. No hay efectos de interacción importantes.

c. Los principales efectos del factor *B* y el factor *C* son importantes.

d. $w = 1.89$; sólo las máquinas 2 y 4 no difieren de modo importante entre sí.

31. La columna de valor *P* muestra que varios efectos de interacción son importantes al nivel .01.

33.

Causa	gl	SC	CM	*f*
A	6	67.32	11.02	
B	6	51.06	8.51	
C	6	5.43	.91	.61
Error	30	44.26	1.48	
Total	48	168.07		

$F_{.05,6,30} = 2.42$, $f_C = .61$, así que H_{0C} no es rechazada.

35.

Causa	gl	SC	CM	*f*
A	4	28.88	7.22	10.7
B	4	23.70	5.93	8.79
C	4	.62	.155	<1
Error	12	8.10	.675	
Total	24	61.30		

Como $F_{.05,4,12} = 3.26$, *A* y *B* son importantes.

37.

Causa	gl	CM	*f*
A	2	2207.329	2259*
B	1	47.255	48.4*
C	2	491.783	503*
D	1	.044	<1
AB	2	15.303	15.7*
AC	4	275.446	282*
AD	2	.470	<1
BC	2	2.141	2.19
BD	1	.273	<1
CD	2	.247	<1
ABC	4	3.714	3.80
ABD	2	4.072	4.17*
ACD	4	.767	<1
BCD	2	.280	<1
ABCD	4	.347	<1
Error	36	.977	
Total	71	93.621	

*Denota una razón *F* importante.

39. a. $\hat{\beta}_1 = 54.38$, $\hat{\gamma}_{11}^{AC} = -2.21$, $\hat{\gamma}_{21}^{AC} = 2.21$.

b.

Causa	Contraste de efecto	CM	*f*
A	1307	71,177.04	436.7
B	1305	70,959.34	435.4
C	529	11,660.04	71.54
AB	199	1650.04	10.12
AC	−53	117.04	<1
BC	57	135.38	<1
ABC	27	30.38	<1
Error		162.98	

41.

Causa	SC	*f*
A	136,640.02	1007.6
B	139,644.19	1029.8
C	24,616.02	181.5
D	20,377.52	150.3
AB	2173.52	16.0
AC	2.52	<1
AD	58.52	<1
BC	165.02	1.2
BD	9.19	<1
CD	17.52	<1
ABC	42.19	<1
ABD	117.19	<1
ACD	188.02	1.4
BCD	13.02	<1
ABCD	204.19	1.5
Error	4339.33	
Total	328,607.98	

$F_{.05,1.32} \approx 4.15$, de modo que sólo los cuatro efectos principales y la interacción *AB* parecen importantes.

43.

Causa	gl	SC	*f*
A	1	.436	<1
B	1	.099	<1
C	1	.109	<1
D	1	414.12	851
AB	1	.003	<1
AC	1	.078	<1
AD	1	.017	<1
BC	1	1.404	3.62
BD	1	.456	<1
CD	1	2.190	4.50
Error	5	2.434	

$F_{.05,1,5} = 6.61$, de modo que sólo el efecto principal del factor *D* se considera importante.

45. a. 1: (1), *ab, cd, abcd;* 2: *a, b, acd, bcd;* 3: *c, d, abc, abd;* 4: *ac, bc, ad, bd.*

b.

Causa	gl	SC	*f*
A	1	12,403.125	27.18
B	1	92,235.125	202.13
C	1	3.125	0.01
D	1	60.500	0.13
AC	1	10.125	0.02
AD	1	91.125	0.20
BC	1	50.000	0.11
BD	1	420.500	0.92
ABC	1	3.125	0.01
ABD	1	0.500	0.00
ACD	1	200.000	0.44
BCD	1	2.000	0.00
Blocks	7	898.875	0.28
Error	12	5475.750	
Total	31	111,853.875	

$F_{.01,1,12} = 9.33$, así que sólo los principales efectos de *A* y *B* son importantes.

47. a. *ABFG;* (1), *ab, cd, ce, de, fg, acf, adf, adg, aef, acg, aeg, bcg, bcf, bdf, bdg, bef, beg, abcd, abce, abde, abfg, cdfg, cefg, defg, acdef, acdeg, bcdef, bcdeg, abcdfg, abcefg, abdefg.* {*A, BCDE, ACDEFG, BFG*}, {*B, ACDE, BCDEFG, AFG*}, {*C, ABDE, DEFG, ABCFG*}, {*D, ABCE, CEFG, ABDFG*}, {*E, ABCD, CDFG, ABEFG*}, {*F, ABCDEF, CDEG, ABG*}, {*G, ABCDEG, CDEF, ABF*} . **b.** 1: (1), *aef, beg, abcd, abfg, cdfg, acdeg, bcdef;* 2: *ab, cd, fg, aeg, bef, acdef, bcdeg, abcdfg;* 3: *de, acg, adf, bcf, bdg, abce, cefg, abdefg;* 4: *ce, acf, adg, bcg, bdf, abde, defg, abcefg.*

49. SSA = 2.250, SSB = 7.840, SSC = .360, SSD = 52.563, SSE = 10.240, SSAB = 1.563, SSAC = 7.563, SSAD = .090, SSAE = 4.203, SSBC = 2.103, SSBD = .010, SSBE = .123, SSCD = .010, SSCE = .063, SSDE = 4.840. Error SC = suma de las SC de dos factores = 20.568, Error CM = 2.057, $F_{.01,1,10} = 10.04$, así que sólo el efecto principal de *D* es importante.

51.

Causa	gl	SC	CM	*f*
Efectos principales de *A*	1	322.667	322.667	980.38
Efectos principales de *B*	3	35.623	11.874	36.08
Interacción	3	8.557	2.852	8.67
Error	16	5.266	0.329	
Total	23	372.113		

$F_{.05,3.16} = 3.24$, por lo que parecen estar presentes interacciones.

53.

Causa	gl	SC	CM	*f*
A	1	30.25	30.25	6.72
B	1	144.00	144.00	32.00
C	1	12.25	12.25	2.72
AB	1	1122.25	1122.25	249.39
AC	1	1.00	1.00	.22
BC	1	12.25	12.25	2.72
ABC	1	16.00	16.00	3.56
Error	4	36.00	4.50	
Total	7			

Sólo el efecto principal para *B* y el efecto de interacción *AB* son importantes en $\alpha = .01$.

55. a. $\hat{\alpha}_1 = 9.00$, $\hat{\beta}_1 = 2.25$, $\hat{\delta}_1 = 17.00$,$\hat{\gamma}_1 = 21.00$, $(\hat{\alpha\beta})_{11} = 0$, $(\hat{\alpha\delta})_{11} = 2.00$, $(\hat{\alpha\gamma})_{11} = 2.75$, $(\hat{\beta\delta})_{11} = .75$, $(\hat{\beta\gamma})_{11} = .50$, $(\hat{\delta\gamma})_{11} = 4.50$

b. Una gráfica de probabilidad normal sugiere que los efectos principales *A, C* y *D* son bastante importantes, y quizá la interacción *CD*. De hecho, agrupando las 4 ss = sc de interacción de tres factores y la SS de interacción de cuatro factores, para obtener una SSE basada en 5 grados de libertad, y luego construyendo una tabla ANOVA sugiere que éstos son los efectos más importantes.

57.

Causa	gl	SC	CM	*f*	$F_{.01}$
A	2	67,553	33,777	11.37	5.49
B	2	72,361	36,181	12.18	5.49
C	2	442,111	221,056	74.43	5.49
AB	4	9696	2424	0.82	4.11
AC	4	6213	1553	0.52	4.11
BC	4	34,928	8732	2.94	4.11
ABC	8	33,487	4186	1.41	3.26
Error	27	80,192	2970		
Total	53	746,542			

Todos los efectos principales son significativos, pero no hay interacciones de importancia.

59. Con base en los valores *P* de la tabla ANOVA, los factores estadísticamente significativos al nivel $\alpha = .01$ son el tipo de adhesivo y el tiempo de cura. El material del conductor no tiene un efecto estadísticamente significativo en la resistencia a la adherencia. No hay más interacciones significativas.

61.

Causa	gl	SC	CM	*f*
A	4	285.76	71.44	.594
B	4	227.76	56.94	.473
C	4	2867.76	716.94	5.958
D	4	5536.56	1384.14	11.502
Error	8	962.72	120.34	$F_{.05,4.8} = 3.84$
Total	24			

H_{0A} y H_{0B} no pueden ser rechazadas, mientras que H_{0C} y H_{0D} son rechazadas.

Capítulo 12

1. a. Las tablas siguientes están basadas en la repetición, cinco veces, de cada uno de los valores de tallo (una vez para las hojas 0 y 1, una segunda vez para las hojas 2 y 3, etc.).

17	0	
17	2 3	
17	4 4 5	
17	6 7	
17		tallo: centenas y decenas
18	0 0 0 0 0 1 1	hoja: unidades
18	2 2 2 2	
18	4 4 5	
18	6	
18	8	

No hay valores atípicos, ni brechas significativas y la distribución casi tiene forma de campana con un grado razonablemente alto de concentración alrededor de su centro, en aproximadamente 180.

0	8 8 9	
1	0 0 0 0	
1	3	
1	4 4 4 4	
1	6 6	
1	8 8 8 9	tallo: unidades
2	1 1	hoja: decenas
2		
2	5	
2	6	
2		
3	0 0	

Un valor típico es alrededor de 1.6, y hay una cantidad razonable de dispersión alrededor de este valor. La distribución es un tanto asimétrica hacia valores grandes, los dos más grandes de los cuales pueden ser candidatos para ser valores apartados.
b. No, porque las observaciones con valores x idénticos tienen valores y diferentes.
c. No, porque los puntos no parecen caer en absoluto cerca de una recta o curva simple.

3. Sí. Sí.

5. b. Sí.
c. Parece haber una relación cuadrática aproximada (los puntos caen cerca de una parábola).

7. a. 5050 **b.** 1.3 **c.** 130 **d.** −130

9. a. .095 **b.** −.475 **c.** .830, 1.305 **d.** .4207, .3446 **e.** .0036

11. a. −.01, −.10 **b.** 3.00, 2.50 **c.** .3627 **d.** .4641

13. a. Sí, porque $r^2 = .972$

15. a.

2	9
3	3 3 5 5 6 6 6 7 7 8 8 9
4	1 2 2 3 5 6 6 8 9
5	1
6	2 9
7	9
8	0

Valor típico en los 40 bajos, cantidad razonable de variabilidad, asimetría positiva, dos valores apartados potenciales.
b. No
c. $y = 3.2925 + .10748x = 7.59$. No; riesgo de extrapolación
d. 18.736, 71.605, .738, sí

17. a. $118.91 - .905x$; Sí. **b.** Estimamos que la disminución esperada en la porosidad asociada con un incremento de 1-pcf por unidad de peso es .905%. **c.** Predicción negativa, ya que y no puede ser negativa. **d.** −.52, .49 **e.** .938, aproximadamente el tamaño de una desviación típica de una recta de regresión. **f.** .974

19. a. $y = -45.5519 + 1.7114x$ **b.** 339.51
c. −85.57 **d.** Las $\hat{y}_i$ son 125.6, 168.4, 168.4, 211.1, 211.1, 296.7, 296.7, 382.3, 382.3, 467.9, 467.9, 553.4, 639.0, 639.0; una recta de 45° que pasa por (0, 0).

21. a. Sí; $r^2 = .985$ **b.** 368.89 **c.** 368.89

23. a. 16,213.64; 16,205.45
b. 414,235.71; sí, porque $r^2 = .961$.

27. $\hat{\beta}_1 = \Sigma x_i Y_i / \Sigma x_i^2$

29.

Conjunto de datos	r^2	s	
1	.43	4.03	Más eficaz: conjunto 3
2	.99	4.03	Menos eficaz: conjunto 1
3	.99	1.90	

31. a. .893 **b.** .01837 **c.** (−.216, −.136)

33. a. (.081, .133)
b. H_a: $\beta_1 > .1$, valor $P = .277$, no

35. a. (.63, 2.44) es un intervalo de confianza de 95%.
b. Sí. $t \approx 3.6$, valor $P \approx .004$
c. No; extrapolación
d. (.54, 2.82), no

37. a. Sí. $t = 7.99$, valor $P \approx 0$. *Nota:* Hay un valor atípico moderado, de modo que la gráfica de probabilidad normal resultante no es del todo satisfactoria.
b. Sí. $t = -5.8$, valor $P \approx 0$, rechace H_0: $\beta_1 = 1$ a favor de H_a: $\beta_1 < 1$

39. $f = 71.97$, $s_{\hat{\beta}_1} = .004837$, $t = 8.48$, valor $P = .000$

43. $d = 1.20$, grado de libertad $= 13$ y $\beta \approx .1$.

45. a. (77.80, 78.38)
b. (76.90, 79.28), mismo centro pero más ancho
c. más ancho, porque 115 está más lejos que $\bar{x}$
d. $t = -11$, valor $P = 0$

47. **a.** IP de 95% es (20.21, 43.69), no
b. (28.53, 51.92), al menos 90%

49. (431.3, 628.5)

51. **a.** 45 es más cercano a $\bar{x} = 45.18$ **b.** (46.28, 46.78)
c. (47.56, 49.84)

53. (a) más angosto que (b), (c) más angosto que (d), (a) más angosto que (c), (b) más angosto que (d)

57. Si, por ejemplo, 16 es la edad mínima de elegibilidad, entonces para casi todos $y \approx x - 16$.

59. **a.** .966
b. El porcentaje en peso de fibra seca para el primer espécimen tiende a ser más grande que para el segundo.
c. Sin cambio **d.** 93.3%
e. $t = 14.9$, valor $P \approx 0$, de modo que parece haber esa relación.

61. **a.** $r = .748$, $t = 3.9$, valor $P = .001$. Usando ya sea $\alpha = .05$ o .01, sí.
b. .560 (56%), igual

63. $r = .773$, pero $t = 2.44 < 2.776$; así que H_0: $\rho = 0$ no puede ser rechazada.

65. **a.** .481

b. $t = 1.98$, valor $P = .07$, por lo que a nivel de .01, no hay asociación lineal. **c.** En el nivel de .01, no hay asociación lineal positiva, pero a nivel de .05, la asociación lineal parece ser positiva.

67. **a.** Rechazar H_0
b. No. Valor $P = .00032 \Rightarrow z \approx 3.6 \Rightarrow r \approx .16$, que indica sólo una relación débil.
c. Sí, pero muy grande $n \Rightarrow \rho \approx .022$, y no hay significación práctica.

69. **a.** Un intervalo de confianza de 95%: (.888, 1.086)
b. Un intervalo de confianza de 95%: (47.730, 49.172)
c. IP de 95%: (45.378, 51.524)
d. Más angosto para $x = 25$ por tanto 25 es más cercano a $\bar{x}$
e. .981

71. **a.** $t = -1.24 > -2.201$, y no se rechaza H_0
b. .970

73. **a.** .507 **b.** .712 **c.** Valor $P = .0013 < .01 = \alpha$, y rechace H_0: $\beta_1 = 0$ y concluya que hay una relación lineal útil.
d. Un intervalo de confianza de 95% es (1.056, 1.275).
e. 1.0143, .2143

75. **a.** $y = 1.69 + .0805x$ **b.** $y = -20.40 + 12.2254x$ **c.** .984 para ambas regresiones.

77. **a.** Una relación lineal importante
b. $y = -.08259 + .044649x$
c. 98.3%
d. .7702, −.0902 **e.** Sí; $t = 19.96$
f. (.0394, .0499) **g.** (.762, .858)

81. **b.** .573

87. $t = -1.14$, y es posible que $\beta_1 = \gamma_1$.

Capítulo 13

1. **a.** 6.32, 8.37, 8.94, 8.37, y 6.32 **b.** 7.87, 8.49, 8.83, 8.94, y 2.83 **c.** Es probable que la desviación sea mucho menor para los valores x del inciso (b).

3. **a.** Sí. **b.** −.31, −.31, .48, 1.23, −1.15, .35, −.10, −1.39, .82, −.16, .62, .09, 1.17, −1.50, .96, .02, .65, −2.16, −.79, 1.74. Aquí e/e^* varía entre .57 y .65, así que e^* es cercano a e/s.
c. No.

5. **a.** Alrededor de 98% de variación observada en grosor está explicada por la relación.
b. Una relación no lineal.

7. **a.** .776 **b.** quizá no, debido a la curvatura.
c. Curvatura sustancial en lugar de un patrón lineal implica que el modelo lineal es inadecuado. Una parábola (regresión cuadrática) proporciona un significativo mejor ajuste

9. Para el conjunto 1, la regresión lineal simple es apropiada. Una regresión cuadrática es razonable para el conjunto 2. En el conjunto 3, (13, 12.74) parece muy inconsistente con la información restante. La pendiente estimada para el conjunto 4 depende principalmente de la observación simple (19, 12.5), y la evidencia para una relación lineal no es obligatoria.

11. **c.** $V(\hat{Y}_i)$ aumenta, y $V(Y_i - \hat{Y}_i)$ disminuye.

13. t con $n - 2$ grados de libertad; .02

15. **a.** Un patrón curvo **b.** Un patrón lineal
c. $Y = \alpha x^{\beta} \cdot \epsilon$ **d.** Un IP de 95% es (3.06, 6.50).
e. Un residuo estandarizado, correspondiente a la tercera observación, es un poco grande. Sólo hay dos residuos estandarizados positivos, pero otros dos son 0 en esencia. Los patrones de una gráfica de residuo estandarizado y una gráfica de probabilidad normal son marginalmente aceptables.

17. **a.** $\sum x_i' = 15.501$, $\sum y_i' = 13.352$, $\sum (x_i')^2 = 20.228$, $\sum x_i' y_i' = 18.109$, $\sum (y_i')^2 = 16.572$, $\hat{\beta}_1 = 1.254$,
$\hat{\beta}_0 = -.468$, $\hat{\alpha} = .626$, $\hat{\beta} = 1.254$ **c.** $t = -1.07$, y no rechaza H_0. **d.** H_0: $\beta = 1$, $t = -4.30$, así se rechaza H_0.

19. **a.** No **b.** $Y' = \beta_0 + \beta_1 \cdot (1/t) + \epsilon'$, donde $Y' = \ln(Y)$,
y entonces $Y = \alpha e^{\beta/t} \cdot \epsilon$. **c.** $\hat{\beta} = \hat{\beta}_1 = 3735.45$, $\hat{\beta}_0 = -10.2045$ $\hat{\alpha} = (3.70034) \cdot (10^{-5})$, $\hat{y}' = 6.7748$, $\hat{y} = 875.5$
d. SSE = 1.39587, SSPE = 1.36594 (usando valores transformados), $f = .33 < 8.68 = F_{.01,1,15}$, y no se rechaza H_0.

21. **a.** $\hat{\mu}_{Y \cdot x} = 18.14 - 1485/x$ **b.** $\hat{y} = 15.17$

23. Para un modelo exponencial, $V(Y|x) = \alpha^2 e^{2\beta x}\sigma^2$, que depende de x. Un resultado similar se cumple para el modelo de potencia.

25. La razón z para β_1 es muy importante, indica que la probabilidad de que un nivel sea aceptable disminuye conforme el nivel aumenta. Estimamos que para cada incremento de 1 dBA en el nivel de ruido, la probabilidad de aceptación disminuye por un factor de .70.

27. **b.** 52.88, .12 **c.** .895 **d.** No **e.** (48.54, 57.22) **f.** (42.85, 62.91)

29. **a.** SSE = 16.8, $s = 2.048$ **b.** $R^2 = .995$ **c.** Sí. $t = -6.55$, valor $P = .003$ (de Minitab) **d.** 98% de niveles de confianza individuales $\Rightarrow$ nivel de confianza conjunta $\geq$ 96%: (.671,3.706), (−.00498, −.00135) **e.** (69.531, 76.186), (66.271, 79.446), usando software

31. **a.** .980
b. .747, mucho menor que .977 para el modelo cúbico.
c. Sí, ya que $t = 14.18$,valor $P = 0$.
d. (6.31, 6.57), (6.06, 6.81)
e. $t = -5.6$, valor $P = 0$

33. **a.** .9671, .9407
b. $.0000492x^3 - .000446058x^2 + .007290688x + .96034944$
c. $t = 2 < 3.182 = t_{.025,3}$, de modo que el término cúbico debe eliminarse. **d.** Idénticos
e. .987, .994, sí

35. $\hat{y} = 7.6883e^{.1799x - .0022x^2}$

37. **a.** 4.9 **b.** Cuando el número de entregas se mantiene fijo, el promedio de cambio en tiempo de viaje asociado con un aumento de 1 milla en distancia recorrida es de .060 hora. Cuando la distancia recorrida se mantiene fija, el promedio de cambio en tiempo de viaje asociado con una entrega extra es de .900 hora. **c.** .9861

39. **a.** 77.3 **b.** 40.4

41. $f = 24.4 > 5.12 = F_{.001,6,30}$, y el valor $P < .001$. El modelo escogido parece ser útil.

43. **a.** 48.31, 3.69 **b.** No. Si x_1 aumenta, debe cambiar x_3 o x_2. **c.** Sí, porque $f = 18.924$, valor $P = .001$.
d. Sí, usando $\alpha = .01$, porque $t = 3.496$ y valor $P = .003$.

45. **a.** $f = 87.6$, valor $P = 0$, parece ser una relación lineal útil entre y y al menos uno de los predictores.
b. .935 **c.** (9.095, 11.087)

47. **b.** Valor $P = .000$, así que concluya que el modelo es útil.
c. Valor $P = .034 \leq .05 = \alpha$, y rechace H_0: $\beta_3 = 0$; el % de basura parece dar información adicional útil.
d. (1479.8, 1531.1), precisión razonable
e. Un IP de 95% es (1435.7, 1575.2).

49. **a.** 96.8303, −5.8303 **b.** $f = 14.9 \geq 8.02 = F_{.05,2,9}$, así se rechaza H_0 y se concluye que el modelo es útil. **c.** (78.28, 115.38) **d.** (38.50, 155.16) **e.** (46.91, 140.66) **f.** No. Valor $P = .208$, H_0: $\beta_1 = 0$ no puede ser rechazada.

51. **a.** No **b.** $f = 5.04 \geq 3.69 = F_{.05,5,8}$. Parece haber una relación lineal útil. **c.** 6.16, 3.304, (16.67, 31.91) **d.** $f = 3.44 < 4.07 = F_{.05,3,8}$, así H_0: $\beta_3 = \beta_4 = \beta_5 = 0$ no puede ser rechazada. Los términos cuadráticos pueden borrarse.

55. **a.** La variable dependiente es $\ln(q)$, y los predictores son $x_1 = \ln(a)$ y $x_2 = \ln(b)$; $\hat{\beta} = \hat{\beta}_1 = .9450$, $\hat{\gamma} = \hat{\beta}_2 = .1815$, $\hat{\alpha} = 4.7836$, $\hat{q} = 18.27$. **b.** Ahora regrese $\ln(q)$ contra $x_1 = a$ y $x_2 = b$. **c.** (1.24, 5.78)

57.

k	R^2	adj. R^2	C_k
1	.676	.647	138.2
2	.979	.975	2.7
3	.9819	.976	3.2
4	.9824		4

a. El modelo con $k = 2$ **b.** No

59. El modelo con predictores x_1, x_3, y x_5

61. No. Todos los valores de R^2 son mucho menores que .9.

63. El impacto de estas dos observaciones debe investigarse más a fondo. No enteramente. La eliminación de la observación #6 seguida de rerregresión también debe considerarse.

65. **a.** Las dos distribuciones tienen cantidades similares de variabilidad, son razonablemente simétricas y no contienen valores atípicos. La diferencia principal es que la mediana de los valores de primer orden es de unos 840, en tanto que es de alrededor de 480 para valores no de primer orden. Un intervalo de confianza de 95% t para la diferencia entre medias es (132, 557).
b. $r^2 = .577$ para el modelo de regresión lineal simple, valor P para utilidad de modelo = 0, ¡pero un residuo estandarizado es −4.11! Incluir un indicador para primer orden y no de primer orden no mejora el ajuste, como tampoco incluir un indicador y pronosticadora de interacción.

67. **a.** Cuando el género, peso y ritmo cardiaco se mantienen fijos, estimamos que el cambio promedio en VO_2 máx asociado con un aumento de 1 minuto en tiempo de caminata es de −.0996.
b. Cuando el peso, tiempo de caminata y ritmo cardiaco se mantienen fijos, la estimación del promedio de diferencia entre VO_2 máx para hombres y mujeres es de .6566.
c. 3.669, −.519 **d.** .706
e. $f = 9.0 \geq 4.89 = F_{.01,4,15}$, de manera que parece ser una relación útil.

69. **a.** No. Hay curvatura suficiente en la gráfica de dispersión.
b. La regresión cúbica da $R^2 = .998$ y un intervalo de predicción de 95% de (261.98, 295.62), y la pronosticadora cúbica parece ser importante (valor $P = .001$). Una regresión de y contra $\ln(x)$ tiene $r^2 = .991$, pero hay un residuo estandarizado muy grande y el diagrama residual estandarizado no es satisfactorio.

71. **a.** $R^2 = .802$, $f = 21.03$, valor $P = .000$. El pH es un candidato para eliminación. Nótese que hay un residuo estandarizado extremadamente grande.
b. $R^2 = .920$, R^2 ajustada = .774, $f = 6.29$, valor $P = .002$
c. $f = 1.08$, valor $P > .10$, no rechazar H_0: $\beta_6 = \cdots = \beta_{20} = 0$. El grupo de predictores de segundo orden no parece ser útil.
d. $R^2 = .871$, $f = 28.50$, valor $P = .000$ y ahora los seis predictores son considerados importantes (el máximo valor P para cualquier relación t es .016); la importancia de pH^2 fue ocultada en la prueba de (c). Nótese que hay dos residuos estandarizados más bien grandes.

73. **a.** $f = 1783$, de modo que el modelo parece ser útil.
b. $t = -48.1 \leq -6.689$, por lo que incluso al nivel .001 la pronosticadora cuadrática debe retenerse.
c. No **d.** (21.07, 21.65) **e.** (20.67, 22.05)

75. **a.** $f = 30.8 \geq 9.55 = F_{.01,2,7}$, y el modelo parece ser útil.
b. $t = -7.69$ y valor $P < .001$, y retenga la pronosticadora cuadrática. **c.** (44.01, 47.91)

77. **a.** Al nivel de significancia .05, sí, ya que $f = 4.06$ y valor $P = .029$. **b.** Sí, porque $f = 20.1$ y $F_{.05,3,12} = 3.49$. La prueba reducida completa de F no puede ser utilizada ya que los predictores en este modelo no son un subconjunto de los del inciso (a).

79. Hay varias opciones razonables en cada caso.

81. **a.** $f = 106$, valor $P \approx 0$ **b.** (.014, .068)
c. $t = 5.9$, rechazar H_0: $\beta_4 = 0$, el porcentaje de no blancos parece ser importante. **d.** 99.514, $y - \hat{y} = 3.486$

Capítulo 14

1. **a.** Rechazar H_0. **b.** No rechazar H_0.
c. No rechazar H_0. **d.** No rechazar H_0.

3. Sí, ya que $\chi^2 = 19.6 > 11.344$, valor $P = 0$.

5. $\chi^2 = 6.61 < 14.684 = \chi^2_{.10,9}$, y no rechazar H_0.

7. $\chi^2 = 4.03$ y valor $P > .10$, y no rechazar H_0.

9. **a.** [0, .2231), [.2231, .5108), [.5108, .9163), [.9163, 1.6094), y [1.6094, ∞) **b.** $\chi^2 = 1.25 < \chi^2_{\alpha,4}$ para cualquier α razonable, de modo que la distribución exponencial especificada es bastante posible.

11. **a.** $(-\infty, -.97)$, $[-.97, -.43)$, $[-.43, 0)$, $[0, .43)$, $[.43, .97)$, y $[.97, \infty)$ **b.** $(-\infty, .49806)$, [.49806, .49914), [.49914, .5), [.5, .50086), [.50086, .50194), y [.50194, ∞) **c.** $\chi^2 = 5.53$, $\chi^2 = 5.53$, $\chi^2_{.10,5} =$, y el valor $P > .10$ y la distribución normal especificada es posible.

13. $\hat{p} = .0843$, $\chi^2 = 280.3 > \chi^2_{\alpha,1}$ para cualquier α tabulada, y el modelo da un mal ajuste.

15. La probabilidad es proporcional a $\theta^{233}(1-\theta)^{367}$, de donde $\hat{\theta} = .3883$. Las cantidades esperadas estimadas son 21.00, 53.33, 50.78, 21.50, y 3.41. Combinando las celdas 4 y 5, $\chi^2 = 1.62$, y no rechace H_0.

17. $\hat{\mu} = 3.167$, de la cual $\chi^2 = 103.98 >> \chi^2_{\alpha,k-1} = \chi^2_{\alpha,7}$ para cualquier α tabulada, y la distribución Poisson da un muy mal ajuste.

19. $\hat{\theta}_1 = (2n_1 + n_3 + n_5)/2n = .4275$, $\hat{\theta}_2 = .2750$, $\chi^2 = 29.1$, $\chi^2_{.01,3} = 11.344$, y rechace H_0.

21. Sí. La hipótesis nula de una distribución poblacional normal no se puede rechazar.

23. Minitab da $r = .967$, y como $c_{.10} = .9707$ y $c_{.05} = .9639$, $.05 <$ valor $P < .10$. Usando $\alpha = .05$, la normalidad se considera posible.

25. $\chi^2 = 23.18 \geq 13.277 = \chi^2_{.01,4}$, y H_0 es rechazada. Las proporciones parecen ser diferentes.

27. Sí. $\chi^2 = 44.98$ y valor $P < .001$.

29. **a.** Sí, ya que $\chi^2 = 213.2$, valor $P = 0$. **b.** No a cualquier nivel de significancia, ya que $\chi^2 = 3.1$, valor $P > .10$.

31. **a.** Sí. M: .26, .25, .29, .20: F: .11, .18, .34, .37
b. Rechace H_0 al nivel de significancia .05 ó .01, ya que $\chi^2 = 12.1$ por tanto $.005 <$ valor $P < .01$.

35. $N_{ij.}/n$, $n_k N_{ij.}/n$, 24

37. $\chi^2 = 3.65 < 5.992 = \chi^2_{.05,2}$, y H_0 no se puede rechazar.

39. Sí. $\chi^2 = 131$ y el valor $P < .001$.

41. $\chi^2 = 22.4$ y el valor $P < .001$, y se rechaza la hipótesis nula de independencia.

43. Valor $P = 0$, y se rechaza la hipótesis nula de homogeneidad.

47. **a.** Valor estadístico de prueba = 19.2, valor $P = 0$
b. Evidencia de una relación débil en el mejor de los casos; valor estadístico de prueba $= -2.13$
c. Valor estadístico de prueba $= -.98$, valor $P > .10$
d. Valor estadístico de prueba = 3.3, $.01 <$ valor $P < .05$

49. La combinación de 6 y 7 en una sola categoría y 8 y 9 en otra da una prueba basada en 6 gl para la que $\chi^2 = .92$ y el valor $P > .9$!

Capítulo 15

1. $s_+ = 35$ y $14 < 35 < 64$, y H_0 no puede ser rechazada.

3. $s_+ = 18 \leq 21$, y H_0 es rechazada.

5. Rechazar H_0 si $s_+ \geq 64$ o $s_+ \leq 14$. Como $s_+ = 72$, H_0 es rechazada.

7. $s_+ = 442.5$, $z = 2.89 \geq 1.645$, y rechace H_0.

9.

d	0	2	4	6	8	10	12	14	16	18	20
$p(d)$	$\frac{1}{24}$	$\frac{3}{24}$	$\frac{1}{24}$	$\frac{4}{24}$	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{4}{24}$	$\frac{1}{24}$	$\frac{3}{24}$	$\frac{1}{24}$

11. $w = 37$ y $29 < 37 < 61$, y H_0 no puede ser rechazada.

13. $z = 2.27 < 2.58$, y H_0 no puede ser rechazada. Valor $P \approx .023$

15. $w = 39 \leq 41$, y H_0 es rechazada.

17. $(\bar{x}_{(5)}, \bar{x}_{(32)}) = (11.15, 23.80)$

19. $(-.585, .025)$

21. $(d_{ij(5)}, d_{ij(21)}) = (16, 87)$

23. $k = 14.06 \geq 6.251$, y se rechaza H_0.

25. $k = 9.23 \geq 5.992$, y se rechaza H_0.

27. $f_r = 2.60 < 5.992$, y no se rechaza H_0.

29. $f_r = 9.62 > 7.815 = \chi^2_{.05,3}$, y se rechaza H_0.

31. $(-5.9, -3.8)$

33. **a.** .021 **b.** $c = 14$ da $\alpha = .058$; $y = 12$, y H_0 no puede ser rechazada.

35. $w' = 26 < 27$, y no se rechaza H_0.

Capítulo 16

1. Todos los puntos de la gráfica caen entre los límites de control.

3. .9802, .9512, 53

5. **a.** 1.67, .67 **b.** 1, .67 **c.** $C_{pk} \leq C_p$, $=$ cuando $\mu =$ (USL + LSL)/2

7. **a.** .0301 **b.** .2236 **c.** .9292

9. LCL = 12.20, UCL = 13.70. No.

11. LCL = 94.91, UCL = 98.17. Parece haber un problema en el 22º día.

13. **a.** 200 **b.** 4.78 **c.** 384.62 (mayor), 6.30 (menor)

15. LCL = 12.37, UCL = 13.53

17. **a.** LCL = 0, UCL = 6.48
b. LCL = .48, UCL = 6.60

19. LCL = .045, UCL = 2.484. Sí, porque todos los puntos están dentro de los límites de control.

21. **a.** LCL = .105, UCL = .357
b. Sí, porque .39 > UCL.

23. $\bar{p} > 3/53$

25. LCL = 0, UCL = 10.1

27. Cuando el área = .6, LCL = 0 y UCL = 14.6; cuando el área = .8, LCL = 0 y UCL = 13.4; cuando el área = 1.0, LCL = 0 y UCL = 12.6.

29.

l:	1	2	3	4	5	6	7	8
d_l:	0	.001	.017	0	0	.010	0	0
e_l:	0	0	0	.038	0	0	0	.054
l:	9	10	11	12	13	14	15	
d_l:	0	.024	.003	0	0	0	.005	
e_l:	0	0	0	.015	0	0	0	

No hay señales fuera de control.

31. $n = 5$, $h = .00626$

33. Las probabilidades hipergeométricas (calculadas en una calculadora HP21S) son .9919, .9317, .8182, .6775, .5343, .4047, .2964, .2110, .1464, y .0994, mientras que las correspondientes probabilidades binomiales son .9862, .9216, .8108, .6767, .5405, .4162, .3108, .2260, .1605, y .1117. La aproximación es satisfactoria.

35. .9206, .6767, .4198, .2321, .1183; el plan con $n = 100$, $c = 2$ es preferible.

37. .9981, .5968, y .0688

39. **a.** .010, .018, .024, .027, .027, .025, .022, .018, .014, .011
b. .0477, .0274 **c.** 77.3, 202.1, 418.6, 679.9, 945.1, 1188.8, 1393.6, 1559.3, 1686.1, 1781.6

41. La gráfica $\bar{X}$ basada en las desviaciones estándar muestrales: LCL = 402.42, UCL = 442.20. La gráfica $\bar{X}$ basada en los rangos muestrales: LCL = 402.36, UCL = 442.26. Gráfica S: LCL = .55, UCL = 30.37. Gráfica R: LCL = 0, UCL = 82.75.

43. Gráfica S: LCL = 0, UCL = 2.3020; como $s_{21} = 2.931 >$ UCL, el proceso parece estar fuera de control en este momento. Como está identificada una causa asignable, recalcule límites después de la eliminación: para una gráfica S, LCL = 0 y UCL = 2.0529; para una gráfica $\bar{X}$, LCL = 48.583 y UCL = 51.707. Todos los puntos en ambas gráficas están entre los límites de control.

45. $\bar{\bar{x}} = 430.65$, $s = 24.2905$; para una gráfica S, UCL = 62.43 cuando $n = 3$ y UCL = 55.11 cuando $n = 4$; para una gráfica $\bar{X}$, LCL = 383.16 y UCL = 478.14 cuando $n = 3$, y LCL = 391.09 y UCL = 470.21 cuando $n = 4$.

Glosario de símbolos y abreviaturas

Símbolo/ Abreviatura	Página	Descripción
n	13	tamaño de muestra
x	13	variable en la que se hacen las observaciones
$x_1, x_2, \ldots, x_n$	13	observaciones muestrales en x
$\sum_{i=1}^{n} x_i$	28	suma de $x_1, x_2, \ldots, x_n$
$\bar{x}$	28	media muestral
μ	29	media poblacional
N	29	tamaño de la población cuando la población es finita
$\tilde{x}$	30	mediana muestral
$\tilde{\mu}$	31	mediana poblacional
$\bar{x}_{tr}$	32	media recortada
x/n	33	proporción de la muestra
s^2	36	varianza de la muestra
s	36	desviación estándar de la muestra
σ^2, σ	37	varianza poblacional y desviación estándar
$n - 1$	38	grados de libertad para una sola muestra
S_{xx}	38	suma del cuadrado de las desviaciones de la media muestral
f_s	39	dispersión de los cuartos muestrales
$\mathcal{S}$	51	espacio muestral de un experimento
$A, B, C_1, C_2, \ldots$	52	varios eventos
A'	53	complemento del evento A
$A \cup B$	53	unión de los eventos A y B
$A \cap B$	53	intersección de los eventos A y B
$\varnothing$	54	evento nulo (evento que no contiene resultados)
$P(A)$	56	probabilidad del evento A
N	61	número de resultados igualmente probables
$N(A)$	62	número de resultados en el evento A
n_1, n_2	65	número de maneras de seleccionar el 1er. (2o.) elemento de un par ordenado
$P_{k,n}$	67	número de permutaciones de tamaño k a partir de n entidades distintas
$\binom{n}{k}$	69	número de combinaciones de tamaño k a partir de n entidades distintas
$P(A\|B)$	73	probabilidad condicional de A dado que B ocurre

Símbolo/ Abreviatura	Página	Descripción
rv	93	variable aleatoria
X	93	una variable aleatoria
$X(s)$	93	valor de la va X asociada con el resultado s
x	93	valor particular de va x
$p(x)$	96	distribución de probabilidad (función masa) de una va X discreta
pmf	97	función de masa de probabilidad
$p(x; \alpha)$	100	fmp con parámetro α
$F(x)$	101	función de distribución acumulada de una va
cdf	101, 144	función de distribución acumulada
$a-$	104	valor X más grande posible menor que a
$E(X), \mu_X, \mu$	108, 148	media o valor esperado de la va X
$E[h(X)]$	109, 149	valor esperado de la función $h(X)$
$V(X), \sigma_X^2, \sigma^2$	111, 150	varianza de la va X
σ_X, σ	111, 150	desviación estándar de la va X
S, F	114	éxito/fracaso en un ensayo simple de un experimento binomial
n	114	número de intentos en un experimento binomial
p	114, 125	probabilidad de éxito en un intento sencillo de un experimento binomial o un experimento binomial negativo
$X \sim \text{Bin}(n, p)$	116	la va X tiene una distribución binomial con parámetros n y p
$b(x; n, p)$	117	fmp binomial con parámetros n y p
$B(x; n, p)$	118	función de distribución acumulada de una va binomial
M	123	número de éxitos en una dicotomía poblacional de tamaño N
$h(x; n, M, N)$	123	fmp hipergeométrica con parámetros n, M, y N
r	125	número de éxitos esperados en un experimento binomial negativo
$nb(x; r, p)$	125	fmp binomial negativa con parámetros r y p
μ	128	parámetro de una distribución de Poisson
$p(x; \mu)$	128	fmp de Poisson

Símbolo/ Abreviatura	Página	Descripción
$F(x; \mu)$	130	fda de Poisson
Δt	131	longitud de un intervalo corto de tiempo
$o(\Delta t)$	131	cantidad que se aproxima a 0 más rápido que Δt
α	131	parámetro de proporción de un proceso de Poisson
$\alpha(t)$	131	función de proporción de una tasa variable de un proceso de Poisson
fdp	139	función de densidad de probabilidad
$f(x)$	139	función de densidad de probabilidad de una va X continua
$f(x; A, B)$	140	fdp uniforme en el intervalo $[A, B]$
$F(x)$	144	función de distribución acumulada
$\eta(p)$	147	100p-ésimo percentil de una distribución continua
$\tilde{\mu}$	148	mediana de una distribución continua
$f(x; \mu, \sigma)$	152	fdp de una va normalmente distribuida
$N(\mu, \sigma^2)$	153	distribución normal con parámetros μ y σ^2
Z	153	va normal estándar
curva Z	154	curva normal estándar
$\Phi(z)$	154	fda de una va normal estándar
z_α	156	valor que captura un área de cola superior α bajo la curva z
λ	165	parámetro para una distribución exponencial
$f(x; \lambda)$	165	fdp exponencial
$\Gamma(\alpha)$	167	función gamma
$f(x; \alpha, \beta)$	168	fdp gamma con parámetros α y β
gl	170	grados de libertad
ν	170	número de gl para una distribución chi-cuadrada
$f(x; \alpha, \beta)$	172	fdp de Weibull con parámetros α y β
$f(x; \mu, \sigma)$	174	fdp lognormal con parámetros μ y σ
$f(x; \alpha, \beta, A, B)$	176	fdp beta con parámetros α, β, A, B
θ_1, θ_2	185	parámetros de ubicación y escala
$p(x, y)$	194	fmp conjunta de dos va discretas X y Y
$p_X(x), p_Y(y)$	195	fmp marginales de X y Y, respectivamente,
$f_X(x), f_Y(y)$	197	fdp marginales de X y Y, respectivamente
$p(x_1, \ldots, x_n)$	200	fmp conjunta de las n va $X_1, \ldots, X_n$
$f(x_1, \ldots, x_n)$	200	fdp conjunta de las n va $X_1, \ldots, X_n$
$f_{Y\|X}(y\|x)$	202	fdp condicional de Y dado que $X = x$
$p_{Y\|X}(y\|x)$	202	fmp condicional de Y dado que $X = x$
$E(Y\|X = x)$	203	valor esperado de Y dado que $X = x$
$E[h(X, Y)]$	206	valor esperado de la función $h(X, Y)$
$\text{Cov}(X, Y)$	207	covarianza entre X y Y
$\text{Corr}(X, Y)$, $\rho_{X,Y}$, ρ	209	coeficiente de correlación para X y Y

Símbolo/ Abreviatura	Página	Descripción
$\bar{X}$	214	media muestral vista como una va
S^2	214	varianza muestral vista como una va
TLC	225	teorema del límite central
θ	240	símbolo genérico para un parámetro
$\hat{\theta}$	240	estimado puntual o estimador de θ
MVUE	247	estimador insesgado de mínima varianza (o estimado)
$\hat{\sigma}_{\hat{\theta},s}$	251	desviación estándar estimada de $\hat{\theta}$
$x_1^*, \ldots, x_n^*$	251	muestra bootstrap
$\hat{\theta}^*$	251	estimación de θ a partir de una muestra bootstrap
mle	259	estimación de máxima probabilidad (o estimador)
IC	270	intervalo de confianza
$100(1 - \alpha)\%$	272	intervalo de confianza para un IC
T	286	variable que tiene una distribución t
ν	286	parámetro de grados de libertad (gl) para una distribución t
t_ν	286	distribución t con ν gl
$t_{\alpha,\nu}$	287	valor que captura un área de cola superior α bajo la curva de densidad t_ν
IP	290	intervalo de predicción
$\chi^2_{\alpha,\nu}$	294	valor que captura un área de cola superior α bajo la curva de densidad chi-cuadrada con ν gl
H_0	301	hipótesis nula
H_a	301	hipótesis alternativa
α	304	probabilidad de un error tipo I
β	304	probabilidad de un error tipo II
μ_0	310	valor nulo en una prueba respecto a μ
Z	310	estadístico de prueba basado en una distribución normal estándar
μ'	313	valor alternativo de μ en el cálculo de β
$\beta(\mu')$	313	probabilidad de error tipo II cuando $\mu = \mu'$
T	316	estadístico de prueba basado en una distribución t
θ_0	323	valor nulo en una prueba respecto a θ
p_0	324	valor nulo en una prueba respecto a p
p'	325	valor alternativo de p en el cálculo de β
$\beta(p')$	325	probabilidad de error tipo II cuando $p = p'$
Ω_0, Ω_a	341	conjuntos disjuntos de valores de parámetros en una prueba de razón de verosimilitud
m, n	346	tamaños de muestra en problemas de dos muestras
Δ_0	347	valor nulo en una prueba respecto a $\mu_1 - \mu_2$
Δ'	350	valor alternativo de $\mu_1 - \mu_2$ en el cálculo de β

Símbolo/ Abreviatura	Página	Descripción
S_p^2	361	estimador agrupado de σ^2
D_i	366	diferencia $X_i - Y_i$ para el par (X_i, Y_i)
$\bar{d}$, s_D	368	diferencia de la media muestral, desviación estándar muestral-de las diferencias para parejas de datos
p	376	valor común de p_1 y p_2 cuando $p_1 = p_2$
F	382	va que tiene una distribución F
ν_1, ν_2	382	grado de libertad para el numerador y el denominador de una distribución F
F_{α,ν_1,ν_2}	382	valor que captura un área de cola superior α bajo una curva F con ν_1, ν_2 gl
ANOVA	391	análisis de varianza
I	392	número de poblaciones en un factor unitario ANOVA
J	394	tamaño común de muestra cuando tamaños de muestra son iguales
X_{ij}, x_{ij}	394	j-ésima observación en una muestra a partir de la i-ésima población
$\overline{X}_{i\cdot}$	394	media de las observaciones en una muestra a partir de la i-ésima población
$\overline{X}_{\cdot\cdot}$	394	media de todas las observaciones de un conjunto de datos
MSTr	395	media cuadrática de los tratamientos
MSE	395	media cuadrática del error
F	397	estadístico de prueba basado en una distribución F
$x_{i\cdot}$	397	total de observaciones en la i-ésima muestra
$x_{\cdot\cdot}$	397	gran total todas las observaciones
SST	398	suma total de cuadrados
SSTr	398	suma del cuadrado del tratamiento
SSE	398	suma del cuadrado de los errores
m, ν	402	parámetros para una distribución de rango estudentizado
$Q_{\alpha,m,\nu}$	402	valor que captura un área de cola superior bajo una curva de densidad asociada al rango estudentizado
μ	408	promedio de la población representada en un factor único ANOVA
$\alpha_1, \ldots, \alpha_I$	408	efectos de tratamiento en un factor único ANOVA
ϵ_{ij}	408	desviación de X_{ij} del valor de su media
$J_1, \ldots, J_I$	412	tamaño de muestra individual en un factor único ANOVA
n	412	número total de observaciones de un factor único ANOVA para un conjunto de datos
$A_1, \ldots, A_I$	414	efectos aleatorios en un factor único ANOVA

Símbolo/ Abreviatura	Página	Descripción
A, B	420	factores en un factor doble ANOVA
K_{ij}	420	número de observaciones cuando un factor A está en el nivel i y el factor B está en el nivel j
I, J	420	número de niveles de factores A y B, respectivamente
$\overline{X}_{i\cdot}$, $\overline{X}_{j}$	421	promedio de observaciones cuando A (B) está en el nivel i (j)
μ_{ij}	421	respuesta esperada cuando A está en el nivel i y B está en el nivel j
α_i, β_j	423	efecto de A (B) en el nivel i (j)
f_A, f_B	424	proporciones F para prueba de hipótesis sobre efectos de los factores
A_i, B_j	430	efecto de los factores en un modelo de efectos aleatorios
σ_A^2, σ_B^2	430	varianzas del efecto de los factores
K	433	tamaño de la muestra para cada par de niveles (i, j)
γ_{ij}	433	interacción entre A y B en los niveles i y j
A_i, B_j, G_{ij}	438	efectos en una combinación o en modelos de efectos aleatorios
α_i, β_j, δ_k	442	efectos principales en un ANOVA trifactorial
γ_{ij}^{AB}, γ_{ik}^{AC}, γ_{jk}^{BC}	442	interacciones de dos factores en un ANOVA trifactorial
γ_{ijk}	442	interacción de tres factores en un ANOVA trifactorial
I, J, K	442–443	número de niveles de A, B, C en un ANOVA trifactorial
β_1, β_0	472	pendiente intersección de una recta de regresión de una población
ε	472	desviación de Y a partir de su valor medio en una regresión lineal simple
σ^2	472	varianza de la desviación aleatoria ϵ
$\mu_{Y\cdot x^*}$	473	valor medio de Y cuando $x = x^*$
$\sigma^2_{Y\cdot x^*}$	473	varianza de Y cuando $x = x^*$
$\hat{\beta}_1$, $\hat{\beta}_0$	478	estimación de mínimos cuadrados de β_1 y β_0
S_{xy}	479	$\Sigma(x_i - \bar{x})(y_i - \bar{y})$
$\hat{y}_i$	481	valor pronosticado de y cuando $x = x_i$
SSE	483	suma de cuadrados del error (residuo)
SST	485	suma total de cuadrados S_{yy}
r^2	485	coeficiente de determinación
$S_{\hat{\beta}_1}$	493	desviación estándar estimada de $\hat{\beta}_1$
r, R	509	coeficiente de correlación muestral
e_i^*	524	residuo estandarizado
$\beta_i (i = 1, \ldots, k)$	543	coeficiente de x_i en una regresión polinomial
$\hat{\beta}_i$	544	estimación por mínimos cuadrados de β_i
R^2	546, 559	coeficiente de determinación múltiple
β_i^*	548	coeficiente en una regresión polinomial centrada

Símbolo/ Abreviatura	Página	Descripción
β_i	553	coeficiente de un predictor x_i de una regresión poblacional
$\hat{\beta}_i$	557	estimación por mínimos cuadrados de β_i
SSE_k, SSE_l	565	SSE para modelos y simplificados completos, respectivamente
Γ_k	579	estimación normalizada total esperada del error
C_k	579	estimación de Γ_k
h_{ii}	582	coeficiente de y_i en $\hat{y}_i$
$\chi^2_{\alpha,\nu}$	596	valor que captura un área de cola superior bajo la curva χ^2 con ν gl
χ^2	597	estadístico de prueba basado en una distribución chi-cuadrada
$p_{10}, \ldots, p_{k0}$	597	valores nulos de una prueba chi-cuadrada de una H_0 sencilla
$\pi_i(\theta)$	603	categoría de probabilidad como una función de los parámetros $\theta_1, \ldots, \theta_m$
I, J	613	número de poblaciones y categorías en cada población cuando se prueba la homogeneidad
I, J	613	número de categorías en cada uno de dos factores cuando se prueba la independencia
n_{ij}	614	número de individuos de una muestra de una población i que caen en la categoría j
$n_{\cdot j}$	614	número total de individuos muestreados en la categoría j
p_{ij}	614	proporción de la población i en la categoría j
$\hat{e}_{ij}$	615	cantidad esperada estimada en la celda i, j
n_{ij}	617	número de muestras que caen en la categoría i del 1er. factor y la categoría j del 2o. factor
p_{ij}	617	proporción de población en la categoría i del 1er. factor y la categoría j del 2o. factor
S_+	627	estadístico de rango con signo
W	635	estadístico de rango con suma
K	645	estadístico de prueba Kruskal-Wallis
R_{ij}	645	rango de X_{ij} entre todas las N observaciones en el conjunto de datos
$\overline{R}_{i\cdot}$	645	promedio de rangos para las observaciones en la muestra a partir de la población o tratamiento i
F_r	647	estadístico de prueba de Friedman
UCL	652	límite superior de control
LCL	652	límite inferior de control
C_p, C_{pk}	653–654	índices de capacidad del proceso
R	658	rango de la muestra
ARL	660	promedio a largo plazo
IQR	661	rango intercuartil
CUSUM	672	suma acumulada
OC	681	característica de operación
AQL	682	nivel aceptable de calidad
LTPD	682	porcentaje de tolerancia en un lote defectuoso
AOQ	685	calidad de salida promedio
AOQL	685	límite de la calidad de salida promedio
ATI	685	promedio del número total de inspecciones

Índice

Este texto líder en el mercado ofrece una amplia introducción a la probabilidad y la estadística para estudiantes de ingeniería y ciencias en todas las especialidades. Demostrado, preciso y alabado por sus excelentes ejemplos, ***Probabilidad y estadística para ingeniería y ciencias*** evidencia la reputación de Jay Devore como un autor destacado y líder en la comunidad académica. Devore hace hincapié en los conceptos, modelos, metodología y aplicaciones en comparación con el desarrollo matemático riguroso y derivaciones. Ayudado por sus ejemplos vivos y realistas, los estudiantes van más allá de simplemente aprender acerca de las estadísticas, conociendo también cómo poner los métodos estadísticos en uso.

Características

- Más de 40 nuevos ejemplos y 100 nuevos problemas han sido investigados y escritos cuidadosamente con los datos reales más actuales.
- El Capítulo 1, "Generalidades y estadística descriptiva", contiene una nueva subsección sobre "El ámbito de la estadística moderna" que describe y ejemplifica cómo la estadística se utiliza en las disciplinas modernas.
- Se ha revisado y simplificado considerablemente el capítulo 8, "Pruebas de hipótesis basadas en una sola muestra", también tiene un nuevo apartado titulado "Más sobre interpretación de los valores *P*".
- Siempre que sea posible, en todo el libro el idioma se ha reforzado y simplificado para mejorar la claridad.

ISBN-13: 978-607-519-536-0
ISBN-10: 607-519-536-X

CENGAGE Learning™

http://latinoamerica.cengage.com