Daniel Peña

FUNDAMENTOS DE ESTADÍSTICA

Alianza Editorial

Daniel Peña

Fundamentos de Estadística

Alianza Editorial

Edición electrónica, 2014
www.alianzaeditorial.es

Juan Ignacio Luca de Tena, 15. 28027 Madrid
ISBN: 978-84-206-8877-0
Edición en versión digital 2014

A Mely, Jorge y Álvaro

Índice

Segunda parte
Modelos

Cuarta parte

Control de calidad

Tablas:

Prólogo

Este libro es el resultado de veinticinco años de experiencia explicando estadística a estudiantes de ingeniería, economía y administración de empresas y otras licenciaturas universitarias. Cubre los conocimientos básicos que estos profesionales deben adquirir como herramientas imprescindibles para su trabajo y como parte de una formación necesaria para entender la ciencia moderna y evaluar la información cuantitativa que como ciudadanos reciben en un mundo donde la estadística juega un papel creciente.

El libro se estructura siguiendo las etapas de construcción de un modelo estadístico. Tras un capítulo introductorio que presenta el contenido global del libro y una breve introducción histórica a los métodos estudiados, los siguientes capítulos siguen la secuencia de una investigación estadística: análisis exploratorio inicial de los datos disponibles (primera parte, datos, capítulos 2 y 3), construcción de un modelo probabilístico (segunda parte, capítulos 4, 5 y 6) y ajuste del modelo a los datos (tercera parte, inferencia, capítulos 7, 8, 9, 10 y 11). Como aplicación de estas ideas, se presenta en la cuarta parte un capítulo de control de calidad, dirigido especialmente a estudiantes que vayan a trabajar en el mundo empresarial, aunque los conceptos y métodos que se exponen son igualmente útiles para mejorar el funcionamiento de cualquier organización.

Este libro está concebido como texto para un primer curso cuatrimestral de estadística orientado a sus aplicaciones. Por esta razón se incluyen temas de gran importancia práctica que no aparecen habitualmente en libros de texto básicos, como la familia Box-Cox de transformaciones, el concepto

de varianza promedio, las relaciones entre los modelos básicos de distribución de probabilidad, las distribuciones mezcladas, el estudio detallado del método de máxima verosimilitud, el concepto de métodos robustos, la combinación de estimadores, la estimación bayesiana, los métodos autosuficientes *(bootstrap)*, los métodos no paramétricos de estimación de densidades, el análisis de homogeneidad de una muestra, el estudio de datos atípicos y la función de autocorrelación muestral. Estas ideas deben introducirse desde el principio porque, de acuerdo con mi experiencia, el estudiante va a necesitarlas en sus primeros análisis estadísticos con datos reales.

A lo largo del libro se ha pretendido ilustrar los conceptos teóricos con ejemplos y, para reforzar y contrastar su asimilación, se han incluido numerosos ejercicios y problemas cuyas soluciones se encuentran al final del volumen. Estos ejercicios se conciben como parte importante del aprendizaje del estudiante y, por tanto, ciertos conceptos teóricos se complementan o generalizan en ellos.

Es tan incompleto estudiar medicina sin ver jamás a un enfermo como estadística sin analizar datos reales. Por otro lado, el análisis de datos hoy es impensable sin utilizar un ordenador. Los ejemplos y análisis de este libro se han realizado con varios programas informáticos, incluyendo Statgraphics, Excel, Minitab, SPSS, S-Plus y Matlab. Cualquiera de estos programas, que se presentan en orden aproximadamente creciente de sofisticación, puede utilizarse para analizar datos estadísticos y es conveniente que el profesor programe las actividades de estudiantes apoyándose en un programa de ordenador que permita explorar las enormes posibilidades del análisis estadístico para comprender realidades complejas y tomar decisiones en incertidumbre. El estudio teórico y la resolución de ejercicios deben completarse con el análisis de problemas reales para que el estudiante compruebe por sí mismo lo que le aporta la teoría estudiada. Por este camino los conceptos teóricos se convierten en herramientas útiles para su futura actividad profesional.

Este libro es una versión revisada del primer tomo de la obra *Estadística: Modelos y Métodos*. La obra se ha revisado, adaptado y reestructurado completamente con tres objetivos. El primero es aprovechar más las posibilidades ofrecidas por la rapidez y simplicidad de los ordenadores actuales. Esto ha llevado en este libro a ampliar la presentación del método de Montecarlo, incluir en el texto con cierto detalle los métodos autosuficientes de estimación *(bootstrap)* mostrando su utilización práctica e introducir numerosos ejercicios y ejemplos que los estudiantes deben resolver utilizando el ordenador. El segundo objetivo es corregir algunos puntos oscuros y mejorar la presentación del material. Esto ha llevado a subdividir los siete capítulos del libro anterior en los trece actuales, a redactar de nuevo muchas secciones, a ampliar la parte de inferencia bayesiana y a reescribir el capítulo de control de calidad. El tercer objetivo es hacer la obra más flexible para distintas audiencias. Por esta razón el segundo tomo de la obra inicial

se ha subdividido en dos libros independientes, *Regresión y diseño de experimentos* y *Análisis de series temporales*, para facilitar su uso como textos en distintos cursos.

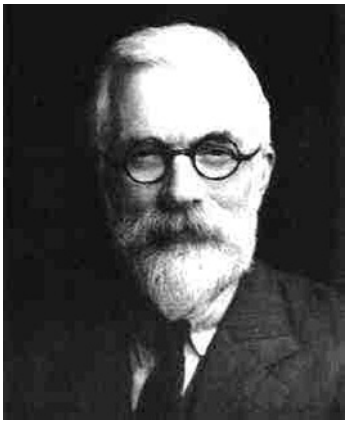
Tengo una deuda especial de gratitud con Rebeca Albacete, María Jesús Sánchez y José Luis Montes, que me han enviado una lista detallada de errores no detectados en ediciones anteriores con excelentes sugerencias de mejora. Gracias a ellos esta edición es más clara y contiene menos erratas. Ángeles Carnero ha conseguido las fotos de estadísticos ilustres buscando en Internet con enorme paciencia y eficacia. Stephan Stigler ha sido de gran ayuda para seleccionar la información histórica. Andrés Alonso, Magdalena Cordero, Pedro Galeano, Miguel Ángel Gómez Villegas, Víctor Guerrero, Jesús Juan, Ana Justel, Agustín Maravall, Francisco Mármol, José Mira, Concepción Molina, Gabriel Palomo, Pilar Poncela, Javier Prieto, Dolores Redondas, Julio Rodríguez, Rosario Romera, Juan Romo, Esther Ruiz, Ismael Sánchez, Santiago Velilla, Teresa Villagarcía, Víctor Yohai y Rubén Zamar han aportado críticas y sugerencias, contribuyendo a mejorar este libro en muchos aspectos. Para todos ellos mi agradecimiento.

Madrid, enero de 2001

En esta nueva edición se han corregido las erratas detectadas y actualizado las referencias. Agradezco mucho la ayuda para llevar a cabo estas mejoras de Adolfo Álvarez, Francisca Blanco, David Casado, Vicente Núñez-Antón, Teresa Villagarcía, Rosario Romera y Henryk Gzyl.

Madrid, junio, 2008

1. Introducción



Ronald Aylmer Fisher (1890-1962)
Científico británico inventor del método de máxima verosimilitud y del diseño estadístico de experimentos. Trabajó en Rothamsted, una estación experimental agrícola en Inglaterra, y fue profesor de eugenesia en la Universidad de Londres. Además de sus numerosas contribuciones a la estadística, que le sitúan como el padre de esta disciplina en el siglo XX, fue un notable genetista, investigador agrario y biólogo.

1.1 La estadística como ciencia

La estadística actual es el resultado de la unión de dos disciplinas que evolucionan independientemente hasta confluir en el siglo XIX: la primera es el cálculo de probabilidades, que nace en el siglo XVII como teoría matemática de los juegos de azar; la segunda es la «estadística» (o ciencia del Estado, del latín *Status*), que estudia la descripción de datos y tiene unas raíces más antiguas. La integración de ambas líneas de pensamiento da lugar a una ciencia que estudia cómo obtener conclusiones de la investigación empírica mediante el uso de modelos matemáticos.

La estadística actúa como disciplina puente entre los modelos matemáticos y los fenómenos reales. Un modelo matemático es una abstracción simplificada de una realidad más compleja, y siempre existirá cierta discrepancia entre lo observado y lo previsto por el modelo. La estadística proporciona

una metodología para evaluar y juzgar estas discrepancias entre la realidad y la teoría. Por tanto, su estudio es básico para todos aquellos que deseen trabajar en ciencia aplicada (sea ésta tecnología, economía o sociología) que requiera el análisis de datos y el diseño de experimentos. La estadística es la «tecnología» del método científico experimental (Mood, 1972).

Además de su papel instrumental, el estudio de la estadística es importante para entender las posibilidades y limitaciones de la investigación experimental, para diferenciar las conclusiones que pueden obtenerse de los datos de las que carecen de base empírica y, en definitiva, para desarrollar un pensamiento crítico y antidogmático ante la realidad.

Muchos ciudadanos ven la estadística con una gran desconfianza: para unos es la ciencia en la que las diferencias individuales quedan ocultas a través de las medias (que se traduce en el dicho popular: «La estadística es la ciencia que explica cómo si tú te comes dos pollos y yo ninguno, nos hemos comido uno cada uno por término medio» y en la famosa frase de Bernard Shaw: «Si un hombre tiene la cabeza en un horno y los pies en una nevera, su cuerpo está a una temperatura media ideal»); para otros es la ciencia mediante la cual con gráficos, tasas de variación y porcentajes se manipula la opinión desde la publicidad, la tecnología o la economía. Vivimos en la era de la estadística y cada aspecto de la actividad humana es medido e interpretado en términos estadísticos.

El único antídoto para esta posible manipulación y para participar efectivamente en la argumentación pública basada en cifras y datos, consustancial a la vida democrática, es comprender el razonamiento estadístico. En este sentido, una formación en los conceptos estadísticos básicos es necesaria para cualquier ciudadano.

1.2 Algunos problemas que resuelve la estadística

Descripción de datos

El primer problema que, históricamente, aborda la estadística es la descripción de datos. Supongamos que se han tomado 1.000 observaciones, que pueden ser gastos de alimentación en una muestra de familias, producción horaria de las máquinas de un taller o preferencias en una muestra de votantes. Se trata de encontrar procedimientos para resumir la información contenida en los datos. Este aspecto se estudia en la primera parte del libro.

Análisis de muestras

Es frecuente que, por razones técnicas o económicas, no sea posible estudiar todos los elementos de una población. Por ejemplo, si para determinar

la resistencia de un elemento es necesario una prueba destructiva, y disponemos de una partida de elementos cuya resistencia se quiere determinar, tendremos que tomar una muestra para no destruir la partida entera. Análogamente, se acude a una muestra para conocer la opinión de la población antes de las elecciones, para estudiar la rentabilidad de un proceso de fabricación o la relación entre el consumo y la renta.

La estadística se utiliza para elegir una muestra representativa y para hacer inferencias respecto a la población a partir de lo observado en la muestra. Éste es el procedimiento aplicado para, por ejemplo:

— Decidir si un proceso industrial funciona o no adecuadamente de acuerdo con las especificaciones.
— Estudiar la relación entre consumo de tabaco y cáncer.
— Juzgar la demanda potencial de un producto mediante un estudio de mercado.
— Orientar la estrategia electoral de un partido político.
— Prever las averías en un taller y diseñar el equipo de mantenimiento.
— Interpretar un test de inteligencia.
— Construir un sistema de reconocimiento de voz.

El análisis de la muestra requiere un modelo probabilístico—cuya construcción será el objeto de la segunda parte de este libro— y la utilización de métodos de inferencia que se expondrán en la tercera parte.

Contrastación de hipótesis

Un objetivo frecuente en la investigación empírica es contrastar una hipótesis. Por ejemplo: ¿Ha mejorado un proceso de fabricación al introducir un elemento nuevo? ¿Es una nueva medicina eficaz para el catarro? ¿Son efectivos el cinturón de seguridad o la limitación de velocidad para reducir las muertes por accidente? ¿Tienen una vida más larga los componentes que tienen el material A que los que no lo tienen? La contrastación de hipótesis requiere una metodología para comparar las predicciones resultantes de la hipótesis con los datos observados y el diseño de un experimento para garantizar que las conclusiones que se extraigan de la experimentación no estén invalidadas por factores no controlados. La metodología estadística para el contraste de hipótesis se expone en el capítulo 10.

Medición de relaciones

Los gastos en alimentación de una familia dependen de sus ingresos, pero es imposible determinar con exactitud cuál será el gasto de una fa-

milia de ingresos dados. Existe entonces una relación no exacta, sino estadística. Determinar y medir estas relaciones es importante porque, debido a los errores de medición, las relaciones que observamos entre variables físicas, sociales o técnicas son, prácticamente siempre, estadísticas.

Preguntas como: ¿Depende la calidad de un producto de los factores A, B y C?, ¿cómo se relaciona el rendimiento escolar con variables familiares y sociológicas?, ¿cuál es la relación entre paro e inflación? tienen que responderse en términos estadísticos. La metodología para analizar estas relaciones se expone en el libro *Regresión y diseño de experimentos,* del mismo autor, que está concebido como extensión de este libro.

Predicción

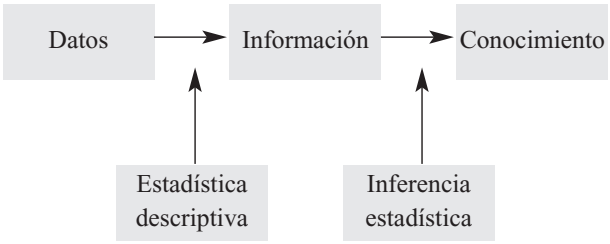
Muchas variables económicas y físicas tienen cierta inercia en su evolución, y aunque sus valores futuros son desconocidos, el estudio de su historia es informativo para prever su evolución futura. Éste es el mecanismo que se utiliza para prever la demanda de un producto, la temperatura en un alto horno o las magnitudes macroeconómicas. La previsión puede mejorarse estudiando la relación entre la variable de interés y otras variables, en el sentido comentado en la sección. Las series temporales se estudian en un texto independiente, concebido como extensión de este libro.

1.3 El método estadístico

El método científico se basa en dos tipos de razonamientos: el deductivo y el inductivo. El método deductivo procede de lo general a lo particular y se utiliza especialmente en el razonamiento matemático: se establecen hipótesis generales que caracterizan un problema y se deducen ciertas propiedades particulares por razonamiento matemático: se establecen hipótesis generales que caracterizan un problema y se deducen ciertas propiedades particulares por razonamientos lógicos. El método inductivo realiza el proceso inverso: a partir de observaciones particulares de ciertos fenómenos se intentan deducir reglas generales.

Una investigación empírica utiliza ambos tipos de razonamiento siguiendo un ciclo deductivo-inductivo: las hipótesis implican propiedades observables en los datos cuyo análisis lleva a formular hipótesis más generales, y así sucesivamente. El método estadístico es el procedimiento mediante el cual se sistematiza y organiza este proceso de aprendizaje iterativo para convertir los datos en información y esta información en conocimiento según el esquema indicado en el cuadro 1.1. La estadística descriptiva se utiliza para sintetizar y resumir los datos transformándolos en información.

Cuadro 1.1 El método estadístico



Esta información es procesada a través de modelos y utilizada para adaptar el modelo a la realidad estudiada, con lo que convertimos la información en conocimiento científico de esa realidad. A continuación se describen las etapas básicas de una investigación estadística.

1.3.1 Planteamiento del problema

Una investigación empírica suele iniciarse con un interrogante del tipo: ¿Cuál es la relación entre...? ¿Qué diferencias existen entre...? ¿Qué ocurriría si...? La primera etapa de la investigación requiere definir el problema en términos precisos, indicando:

a) El ámbito de aplicación, es decir, *la población* que se quiere investigar. Esto exige definir sus límites y caracterizar a sus miembros sin ambigüedad.
b) *Las variables* que debemos observar y cómo medirlas.

Por ejemplo, supongamos que deseamos conocer si la procedencia familiar de un estudiante está relacionada con su rendimiento académico. Tendremos que comenzar definiendo la población que queremos estudiar (por ejemplo, estudiantes matriculados por primera vez en primer curso de una universidad concreta), las variables que definen la procedencia familiar (zona geográfica, estudios de los padres, etc.) y las variables que definen el rendimiento (por ejemplo, nota media en el examen de junio).

Esta fase es fundamental, ya que las conclusiones sólo se aplican a los miembros de la población definida y su validez depende de una selección adecuada de las variables a estudiar.

El resultado de esta fase es una variable respuesta o explicada observable en una o varias poblaciones definidas sin ambigüedad, y un conjunto de variables que podrían explicar esta variable respuesta y que llamaremos variables explicativas.

1.3.2 Construcción de un modelo estadístico

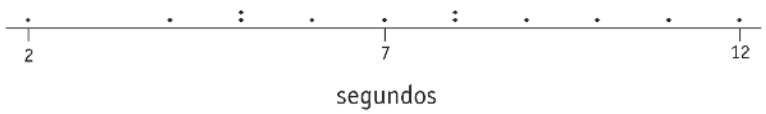
Los modelos estadísticos pueden clasificarse en función de la información que utilizan y del objetivo que pretenden. Cuando la información utilizada corresponde a una única variable, se denominan *modelos univariantes,* cuando incluye además los valores de una o más variables explicativas, se denominan *modelos explicativos.* Por otro lado, si el objetivo es investigar las variables en un instante temporal dado, se denominan *estáticos* o de corte *transversal* (por ejemplo, la relación entre renta y ahorro de las familias españolas en el año 2000), mientras que cuando se desea representar una evolución a lo largo del tiempo se denominan *dinámicos* o *longitudinales.*

En cualquiera de estos cuatro casos, los modelos estadísticos que vamos a estudiar corresponden a una descomposición de los valores de una variable respuesta, *y,* en dos partes. Una parte predecible o sistemática y otra aleatoria, impredecible o residual. El modelo estadístico define la forma de la parte predecible, que representa la respuesta media, y la variabilidad de la impredecible respecto a esa respuesta media. Esta descomposición puede escribirse como:

observación (y) = parte sistemática (predecible)
+ parte aleatoria (impredecible)

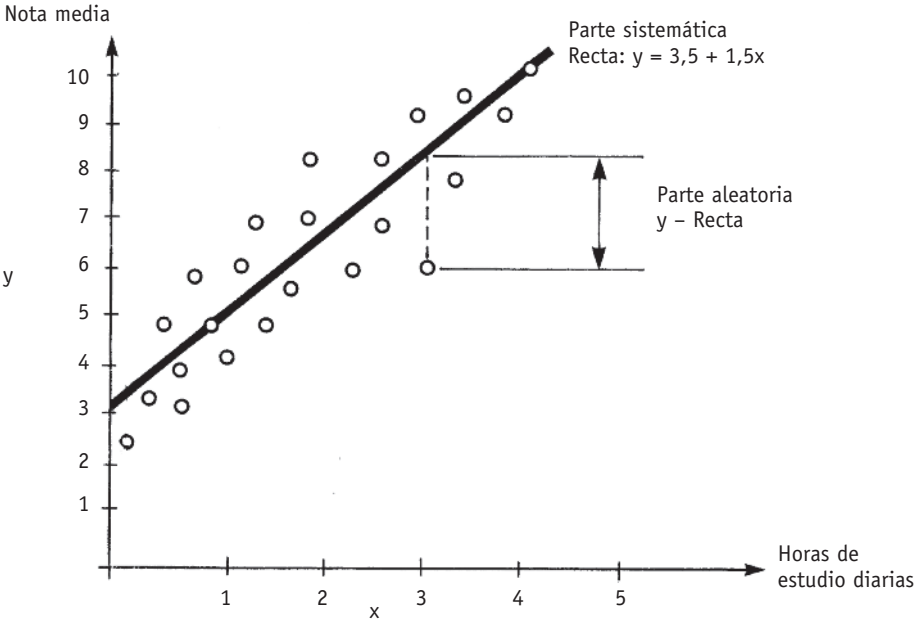
Las figuras 1.1 y 1.2 presentan dos ejemplos de esta descomposición.

Figura 1.1 Tiempo de respuesta en Internet



La primera presenta el tiempo requerido para llegar a una dirección de Internet en doce ocasiones. Cada observación se ha representado por un punto, y la figura muestra que la conexión más rápida se hizo en 2 segundos y la más lenta en 12. Se observa que el tiempo oscila alrededor de un valor central de 7 segundos. Un modelo simple para esta situación es suponer que la conexión se hace en promedio en 7 segundos, pero hay una variabilidad aleatoria en la conexión, de manera que puede tardarse 5 segundos por arriba o por abajo de este valor.

Figura 1.2 Relación entre horas de estudio y nota media



La figura 1.2 representa la relación entre la nota media (variable *y*) obtenida por un grupo de estudiantes en una asignatura y las horas diarias (variable *x*) que en promedio han dedicado a su estudio. Se observa que la nota media depende de las horas de estudio y que los datos se distribuyen alrededor de una recta, que será la parte sistemática o predecible. Esta recta indica un crecimiento lineal de la nota media con el número de horas de estudio. La parte aleatoria será la diferencia entre los valores observados y la recta, y recoge el efecto de todas las variables no consideradas en el modelo (inteligencia de estudiantes, preparación previa, etc.) que producen la variabilidad respecto a la relación promedio.

Estos dos ejemplos son modelos estáticos, ya que estudian la variabilidad en un momento temporal dado. Los modelos de las figuras 1.3 y 1.4 son modelos dinámicos: el primero es *extrapolativo,* ya que utiliza únicamente la información histórica de una serie; el segundo es *explicativo,* ya que introduce otras series como variables explicativas. La figura 1.3 presenta la serie del número de vehículos matriculados cada mes en España en un período de 12 años. La parte sistemática o predecible es ahora mucho más compleja, ya que es la suma de dos componentes:

a) Un componente de tendencia que hace crecer, en promedio, las matriculaciones según una línea recta cuya pendiente varía con el tiempo.

Figura 1.3 Descomposición de la serie de matriculación de vehículos en parte sistemática y parte impredecible o aleatoria

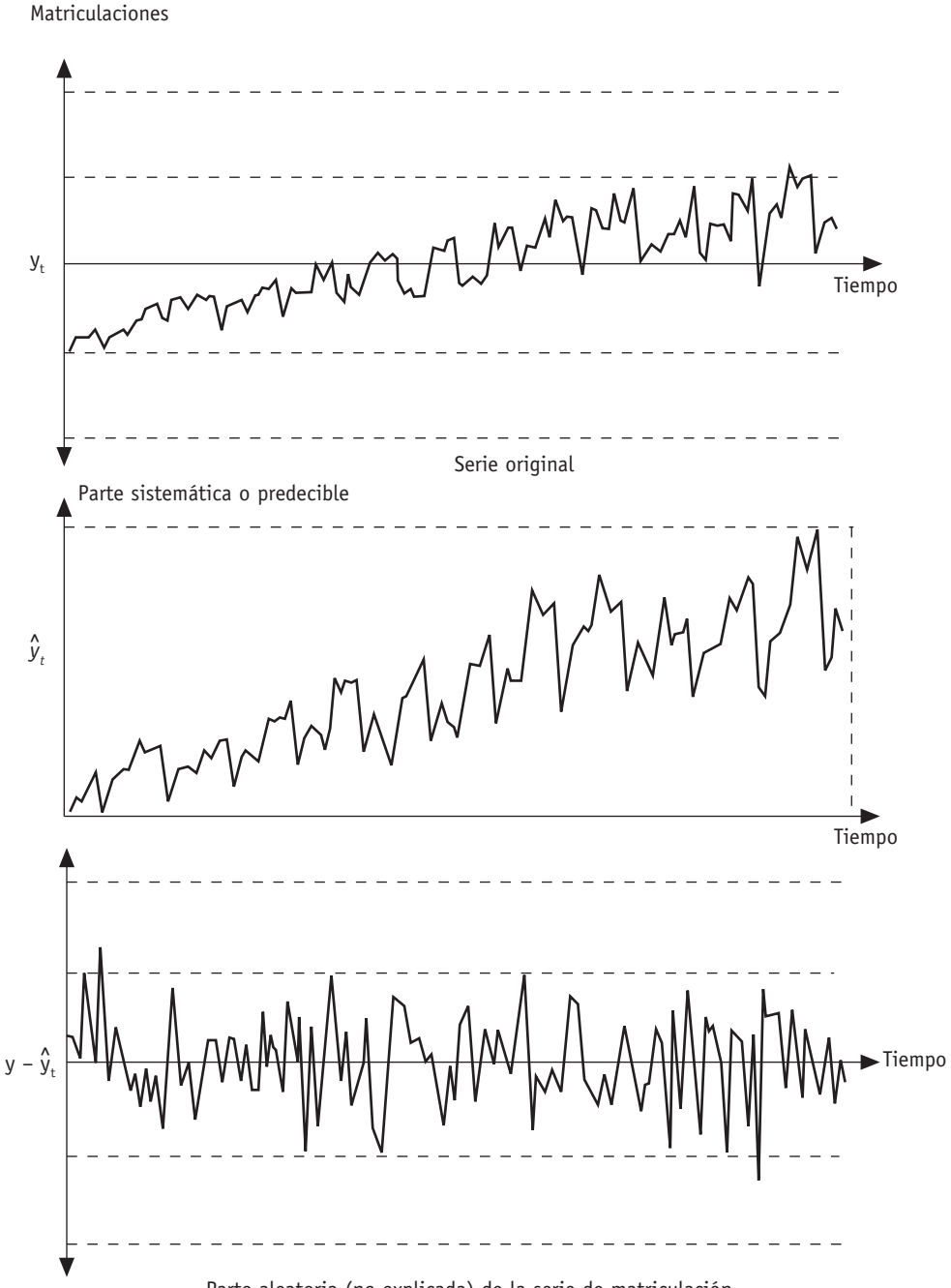
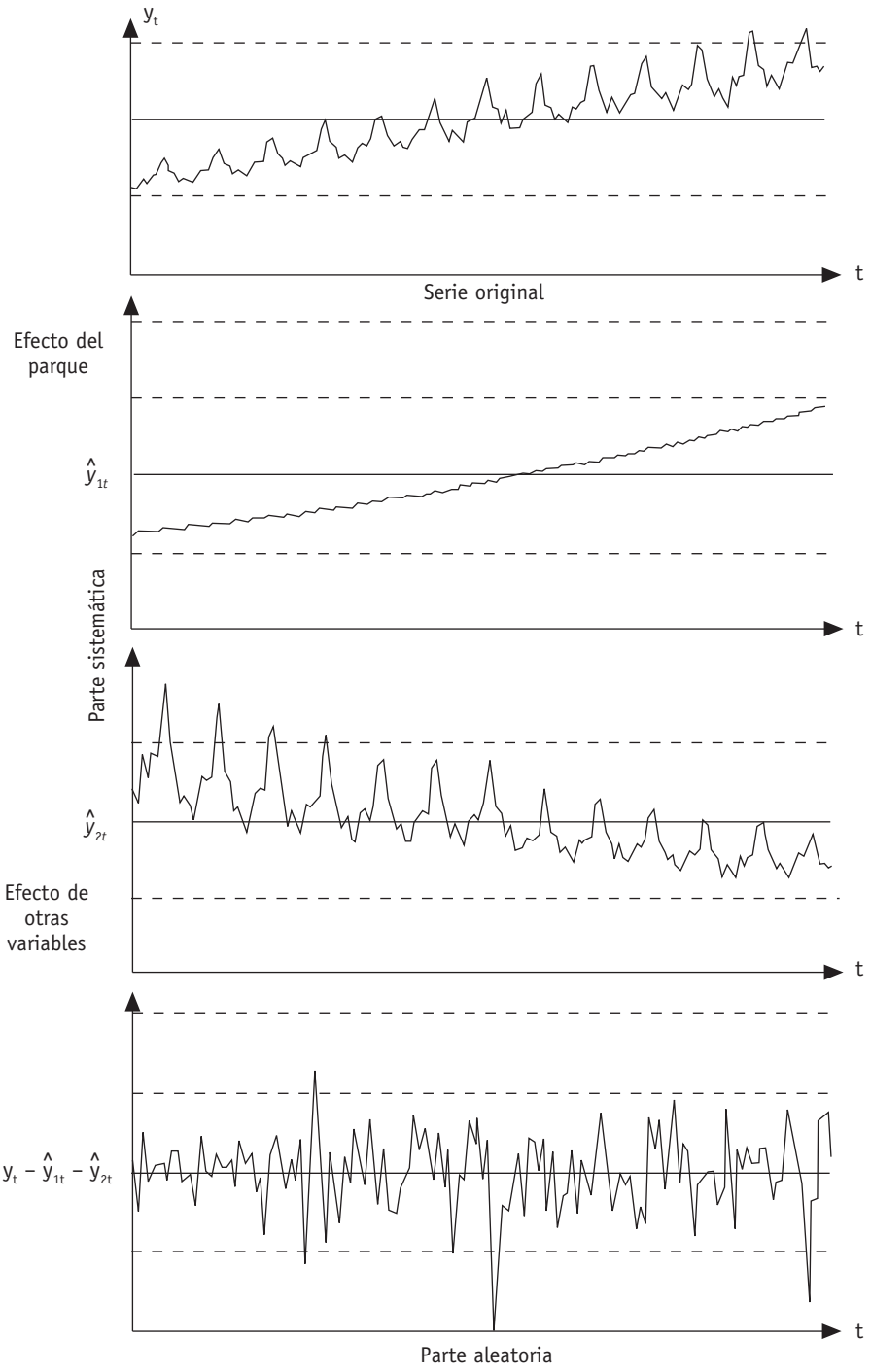


Figura 1.4 Descomposición de la serie de consumo de gasolina



b) Un componente estacional, que hace que cada mes el número de matriculaciones esperadas sea distinto: cada año, descontando la tendencia, se matriculan más vehículos siempre en mayo que en marzo.

Si restamos al número de matriculaciones cada mes, variable y_t, la tendencia y la estacionalidad, cuya suma es la parte predecible, variable $\hat{y}_t$, obtenemos la parte aleatoria o no explicada de la serie que nos proporciona la variabilidad de los datos respecto al valor medio o sistemático de la variable (véase la figura 1.3).

La figura 1.4 presenta un análisis para explicar la evolución de la serie de consumo de gasolina en función del parque de vehículos. La parte sistemática o previsible es ahora la suma de dos componentes. La primera es el efecto debido al aumento del parque de vehículos $\hat{y}_{1t}$, que es una tendencia lineal continuada por el crecimiento del parque. La segunda es la parte sistemática, debida a las otras variables no incluidas en el modelo pero cuya evolución se ha incorporado a la historia de la serie de gasolina. Este efecto es la suma a su vez de una tendencia y de un componente estacional (el consumo de gasolina aumenta en verano) y produce una tendencia decreciente (que puede ser debida al aumento de la eficiencia de los vehículos y a los aumentos del precio de la gasolina) con un efecto estacional superpuesto. Finalmente, la parte aleatoria es la diferencia entre la serie observada y la suma de estos dos componentes explicados, parte explicada o sistemática.

Estos ejemplos muestran las características generales de los modelos estadísticos más frecuentes. Conceptualmente, una variable cualquiera, y, será función de otro gran número de variables, algunas de las cuales pueden no ser observables y cuyo número exacto se desconoce. Un modelo estadístico es una aproximación operativa de esta realidad, que tiene en cuenta explícitamente las variables observables presumiblemente más importantes, y engloba en la parte aleatoria los efectos del resto. Una extensión de estos modelos son los modelos *multivariantes*, donde el interés se centra en un conjunto de variables que se desea explicar conjuntamente. El capítulo 6 presenta una introducción a estos modelos en el caso estático.

1.3.3 Recogida de la información muestral

Una vez construido un modelo del problema, tendremos que medir los valores de las variables de interés. Esta recogida de información puede hacerse de dos formas:

a) Por muestreo.
b) Con un diseño de experimentos.

El muestreo consiste en observar pasivamente una muestra de las variables y anotar sus valores; se utiliza especialmente en modelos extrapolativos.

El diseño de experimentos consiste en fijar los valores de ciertas variables y observar la respuesta de otras. Debe utilizarse siempre que sea posible cuando se desee construir un modelo explicativo. Unicamente tendremos una base empírica sólida para juzgar respecto a relaciones de causalidad entre variables cuando los datos se obtengan mediante un adecuado diseño experimental.

Los fundamentos del muestreo se exponen en el capítulo 7, y los métodos de diseño experimental, en el segundo texto de este trabajo.

1.3.4 Depuración de la muestra

Una regla empírica ampliamente contrastada (Huber, 1984) es esperar entre un 2 y un 5% de observaciones con errores de medición, transcripción, etc. Por tanto, antes de utilizar los datos muestrales conviene aplicar técnicas estadísticas simples, como las que se presentan en el capítulo 2, para identificar valores anómalos y eliminar los errores de medición.

1.3.5 Estimación de los parámetros

Los modelos estadísticos dependen de ciertas constantes desconocidas que llamaremos parámetros. A veces se dispone de información a priori respecto a sus valores, y otras esta información inicial será muy pequeña con relación a la que aportará la muestra. La fase de estimación consiste en utilizar la información disponible para estimar los valores de estos parámetros, así como cuantificar el posible error en la estimación. Los fundamentos de la teoría de la estimación, que son generales para cualquier modelo estadístico, se estudiarán en la tercera parte en los capítulos 7, 8 y 9.

1.3.6 Contrastes de simplificación

Una vez estimados los valores de los parámetros, estudiaremos si el modelo puede simplificarse: por ejemplo, dos parámetros pueden aproximadamente ser iguales, otro puede ser cero, etc. El objetivo de esta fase es conseguir un modelo tan simple como sea posible, es decir, sin más parámetros que los necesarios. Esta fase es especialmente importante en los modelos explicativos, pero aparece en mayor o menor medida en toda investigación estadística. La teoría de contraste de hipótesis se estudiará en el capítulo 10.

1.3.7 Crítica y diagnosis del modelo

Los resultados de las etapas 5 y 6 anteriores se obtienen suponiendo que el modelo es correcto. Esta fase investiga la compatibilidad entre la información empírica y el modelo. De especial interés es comprobar que la parte aleatoria lo es realmente, es decir, no contiene ninguna estructura sistemática. Este aspecto se estudia en el capítulo 12.

Si después de esta fase aceptamos el modelo como correcto, lo utilizaremos para tomar decisiones (capítulo 11) o realizar previsiones de la variable. En caso contrario volveremos a la fase 2 y reformularemos el modelo, repitiendo el proceso hasta conseguir un modelo correcto. Este aspecto cíclico de la investigación se resume en el cuadro 1.2.

La metodología aquí expuesta es iterativa y utiliza tanto el razonamiento deductivo (especialmente en las etapas 2 y 3) como el inductivo (desde la 4 hasta la 7). El cuadro 1.3 presenta dos ejemplos de investigaciones estadísticas: la primera utiliza modelos extrapolativos estáticos y la segunda un modelo explicativo estático.

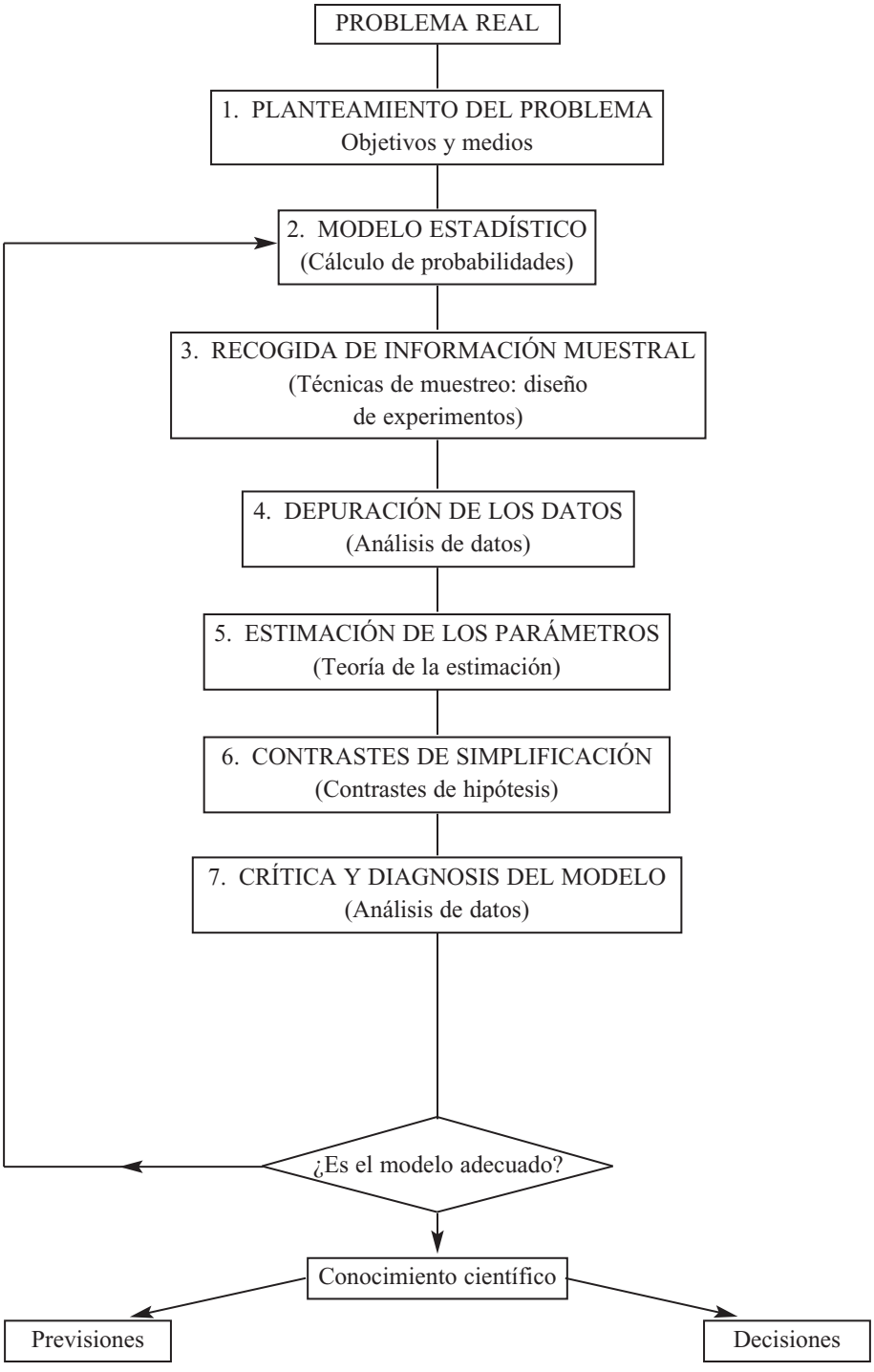
1.4 Notas sobre la historia de la estadística

El conocimiento de la historia de una disciplina es importante, al menos en tres aspectos: el primero, para entender su estado actual de desarrollo y la relación entre sus partes; el segundo, para comprender su terminología, ya que el nombre de una técnica o de un método suele estar asociado a sus orígenes históricos; el tercero, para prever su desarrollo futuro. Por estas razones, consideramos conveniente presentar brevemente algunos rasgos fundamentales de la evolución de la estadística.

1.4.1 El cálculo de probabilidades

La abundante presencia del hueso astrágalo de oveja o ciervo (que constituye el antecedente inmediato del dado) en las excavaciones arqueológicas más antiguas parece confirmar que los juegos de azar tienen una antigüedad de más de 40.000 años, y la utilización del astrágalo en culturas más recientes, Grecia, Egipto y posteriormente Roma, ha sido ampliamente documentada. En las pirámides de Egipto se han encontrado pinturas que muestran juegos de azar que provienen de la primera dinastía (3500 a.C.), y Herodoto se refiere a la popularidad y difusión en su época de los juegos de azar, especialmente mediante la tirada de astrágalos y dados. Los dados más antiguos que se han encontrado se remontan a unos 3.000 años a.C. y se utilizaron tanto en el juego como en ceremonias religiosas.

Cuadro 1.2 Etapas de construcción de un modelo estadístico



Cuadro 1.3 Dos ejemplos de investigaciones estadísticas

Pregunta	¿Cómo diseñar un puesto de servicio?	¿Cómo aumentar el rendimiento de un proceso?
MODELO	Variables: — Número de clientes (x_1) — Tiempo de servicio (x_2) Hipótesis: los clientes • Llegan independientemente. • La probabilidad de llegada de un cliente aumenta exponencialmente con el tiempo. Hipótesis: el tiempo de servicio • Depende de muchos pequeños factores.	Variables: — Rendimiento en % (y) — Temperatura x_1 — Concentración x_2 Hipótesis: • El rendimiento aumenta en promedio linealmente con la temperatura y la concentración. • Para valores fijos de x_1 y x_2 el rendimiento varía aleatoriamente alrededor de su valor medio.
RECOGIDA DE INFORMACIÓN	Muestreo del sistema para estudiar las llegadas de clientes y tiempos de servicio.	Diseño de un experimento en que se varíen x_1 y x_2 y se mida y.
ESTIMACIÓN DE PARÁMETROS	Estimar: • λ, tasa media de llegada. • μ, tiempo medio de servicio. • σ, variabilidad en el tiempo de servicio.	Estimar: • El efecto de la temperatura (b) y el de la concentración (c) sobre el rendimiento. • La variabilidad experimental.
CONTRASTES DE SIMPLIFICACIÓN	¿Tienen todas las semanas la misma λ? ¿Los clientes, el mismo μ y σ?	¿Es el efecto de la temperatura y concentración idéntico ($b = c$)? ¿Puede suponerse $b = 0$?
CRÍTICA DEL MODELO	¿Es cierta la independencia entre llegadas? ¿Son la variabilidad de x_1 y x_2 en la muestra consistentes con las hipótesis?	¿Es la relación entre y (x_1, x_2) lineal? ¿Es la variabilidad de y para x_1, x_2 fijos independiente de los valores concretos de x_1 y x_2?

En las civilizaciones antiguas, el azar se explicaba mediante la voluntad divina. Los oráculos, sacerdotes o pitonisas de Grecia y Roma utilizaban la configuración resultante de tirar cuatro dados para predecir el futuro y revelar la voluntad favorable o desfavorable de los dioses. Por ejemplo, en Grecia clásica y Roma la aparición de la combinación Venus (aparición de 1, 3, 4, 6 al tirar cuatro dados) era favorable, y se ha descubierto en Asia Menor una completa descripción de la interpretación profética de los posibles resultados al tirar cuatro dados. Prácticas similares se han encontrado en culturas tan distantes como la tibetana, la india o la judía.

Como no es posible encontrar una causa o conjunto de causas que permitan predecir el resultado de tirar un dado, las culturas antiguas basadas en el determinismo atribuyeron los resultados de fenómenos aleatorios (dados, presencia de lluvia o fenómenos climáticos, etc.) a la voluntad divina. Piaget ha hecho notar que esta actitud mágica ante el azar se manifiesta igualmente en los niños.

El Renacimiento supuso un nuevo enfoque global de la concepción del mundo, e indujo una observación cualitativamente distinta de muchos fenómenos naturales. En concreto, el abandono progresivo de explicaciones teológicas conduce a una reconsideración de los experimentos aleatorios, y los matemáticos italianos de comienzos del siglo XVI empiezan a interpretar los resultados de experimentos aleatorios simples. Por ejemplo, Cardano, en 1526, establece, por condiciones de simetría, la equiprobabilidad de aparición de las caras de un dado a largo plazo, y Galileo (1564-1642), respondiendo a un jugador que le preguntó por qué es más difícil obtener 9 tirando 3 dados que obtener 10, razonó que de las 216 combinaciones posibles equiprobables 25 conducen a 9 y 27 a 10. Señalamos este dato porque la diferencia empírica entre obtener 9 o 10 es únicamente de $2/216 \simeq 0{,}01$, lo que muestra cómo a finales del siglo XVI existía un intuitivo pero preciso análisis empírico de los resultados aleatorios.

El desarrollo del análisis matemático de los juegos de azar se produce lentamente durante los siglos XVI y XVII, y algunos autores consideran como origen del cálculo de probabilidades la resolución del problema de los puntos en la correspondencia entre Pascal y Fermat en 1654. El problema planteado a estos autores por el caballero de Meré, un jugador empedernido de la Francia del XVII, fue cómo debería repartirse el dinero de las apuestas depositado en la mesa si los jugadores se vieron obligados (presumiblemente por la policía, ya que el juego estaba entonces prohibido) a finalizar la partida sin que existiera un ganador.

El cálculo de probabilidades se consolida como disciplina independiente en el período que transcurre desde la segunda mitad del siglo XVII hasta comienzos del siglo XVIII. En ese período, la teoría se aplica fundamentalmente a los juegos de azar.

Durante el siglo XVIII el cálculo de probabilidades se extiende a problemas físicos y actuariales (seguros marítimos). El factor principal impulsor

de su desarrollo durante este período es el conjunto de problemas de astronomía y física que surgen ligados a la contrastación empírica de la teoría de Newton.

La obra de Newton (1642-1727) constituyó la mayor revolución científica de los siglos XVII y XVIII y su influencia en la evolución de las ciencias físicas es ampliamente conocida. En astronomía, Newton no solamente explicó las leyes de Kepler por el principio de gravitación universal, sino que estableció un modelo global para estudiar las relaciones entre los cuerpos estelares. En física, estableció una teoría común para explicar fenómenos que habían sido objeto de estudios fragmentarios e incompletos como péndulos, planos inclinados, mareas, etc. En matemáticas, contribuyó con Leibnitz a la creación del cálculo diferencial e integral.

Durante el siglo XVIII y parte del XIX la investigación en física y astronomía está dirigida por el paradigma de Newton. Esta investigación se centra en: a) campos de observación y experimentación que la teoría de Newton señala como especialmente relevantes; b) contrastación de las predicciones de la teoría con los datos; c) extender las aplicaciones de la teoría en otros campos. Estas investigaciones van a ser de importancia fundamental en el desarrollo de la estadística.

Un primer problema fue el tratamiento de los errores de medición. Se disponía de varias medidas independientes de una determinada magnitud física y se presentaba el interrogante de cómo combinarlas para obtener un resultado más preciso. Aunque este problema se había planteado en la astronomía desde la antigüedad, la necesidad de comparar con exactitud los datos observados con la teoría requería un tratamiento riguroso del mismo, que va a dar lugar a la teoría de errores.

D. Bernoulli (1700-1782) proporciona la primera solución al problema de estimar una cantidad desconocida a partir de un conjunto de mediciones que, por el error experimental, presentan variabilidad. También desarrolló un test estadístico para determinar si puede aceptarse la hipótesis de que el ordenamiento de las órbitas de los planetas es aleatorio. Este autor fue pionero en la aplicación del cálculo infinitesimal al cálculo de probabilidades.

Pierre Simon, marqués de Laplace (1749-1827), introdujo la primera definición explícita de probabilidad y desarrolló la ley normal como modelo para describir la variabilidad de los errores de medida. También se planteó el problema de predecir una variable conociendo los valores de otras relacionadas con ella y formuló y estimó el primer modelo explicativo estadístico. Es de señalar que, aunque sus procedimientos matemáticos fueron muy «ad hoc», sus resultados fueron sorprendentemente precisos.

La segunda contribución fundamental de este período es debida a Legendre (1752-1833) y Gauss (1777-1855), que resuelven de manera general el problema siguiente de estimación de modelos estáticos: según la teoría, la posición de un planeta en el instante *t,* que llamaremos y_t, es función de

las posiciones de k cuerpos, que representaremos por $x_1, ..., x_k$, y de ciertas constantes desconocidas $\theta_1, ..., \theta_k$. Es decir,

$$y_t = f(\theta_1, ..., \theta_h; x_1, ..., x_k)$$

Disponemos de ciertas observaciones —con cierto error de medida— de las posiciones del planeta y de los cuerpos en cuestión. ¿Cómo determinar las constantes $\theta_1, ..., \theta_h$? ¿Cómo predecir y_t, con la mayor precisión posible dada una observación concreta de valores $x_1, ..., x_k$?

Legendre resolvió estos problemas inventando el método de estimación de mínimos cuadrados, que es todavía hoy la herramienta más utilizada para estimar modelos estadísticos, y Gauss demostró su optimalidad cuando los errores de medida siguen una distribución normal.

Durante la primera mitad del siglo XIX, los matemáticos-astrónomos continúan ampliando la teoría de errores y podemos observar la aparición de problemas y métodos que van a tener gran influencia posterior. Bravais (1846), geólogo y astrónomo, es el primero en considerar la relación entre errores de medida dependientes entre sí, Benjamin Pierce (1852) propone el primer criterio para rechazar observaciones heterogéneas con el resto y S. Newcomb, el más famoso astrónomo americano del XIX, introduce los primeros métodos de estimación cuando hay errores fuertes en algunos datos (estimación robusta).

Por lo tanto, a mediados del siglo XIX existen ya las herramientas básicas que van a dar lugar a la estadística actual. Sin embargo, la aplicación de estos principios va a restringirse a la física y la astronomía, sin ejercer influencia sobre otras áreas de conocimiento.

En particular, estos avances tienen poca influencia sobre una disciplina científica cuyo campo de estudio es el análisis cuantitativo de datos demográficos, sociales y económicos y que se conoce, desde el siglo XVII, con el nombre de estadística.

1.4.2 La estadística hasta el siglo XIX

Desde la antigüedad, los estados han recogido información sobre la población y riqueza que existía en sus dominios. Los censos romanos, los inventarios de Carlomagno de sus posesiones, etc., pueden considerarse precedentes de la institucionalización de la recogida de datos demográficos y económicos por los estados modernos, principalmente por razones fiscales. Esta aritmética política o estadística descriptiva evoluciona durante los siglos XVII y XVIII tomando progresivamente un carácter más cuantitativo.

El primer intento de aplicar un razonamiento propiamente estadístico, en el sentido actual del término, a datos demográficos es debido, en 1662, a Graunt. Este autor se planteó el problema de estimar la población inglesa

de su época y fue capaz, a partir de una muestra, de estimar por primera vez tasas de mortalidad por edades y deducir la frecuencia de nacimientos de hombres y mujeres, entre otros análisis demográficos relevantes. El tipo de razonamiento de Graunt es puramente analítico y desligado completamente del concepto de probabilidad. En la misma línea Petty, en su *Political Arithmetic*, publicado en 1690, analiza datos demográficos, así como datos económicos de ingresos, educación y comercio.

Las primeras tablas completas de mortalidad fueron publicadas por Edmund Halley en 1693, que estudió el problema de los seguros de vida. Durante el siglo XVIII se produce un rápido crecimiento, principalmente en Inglaterra, de los seguros de vida y los seguros marítimos y, debido en gran parte a la influencia de las ideas de Graunt y Petty, se comienzan a realizar los primeros censos oficiales. El primer censo del que se tiene noticias fue realizado por España en Perú en 1548 bajo la dirección del virrey D. Pedro de la Fasca. En Europa, el primer censo se realiza en Irlanda en 1703, y en España, el primero se efectúa en 1787 impulsado por el conde de Floridablanca. A comienzos del siglo XIX puede afirmarse que la casi totalidad de los países europeos recogen información oficial mediante censos de datos demográficos, económicos, climáticos, etc. Paralelamente, surgen las Agencias Oficiales de Estadística y en 1834 se crea en Londres la Royal Statistical Society, seguida, en 1839, por la American Statistical Association.

Durante el siglo XVIII y la mayor parte del siglo XIX, la estadística evoluciona como ciencia separada del cálculo de probabilidades. Aunque A. de Moivre y Deparcieux, entre otros, aplican el cálculo de probabilidades a datos demográficos, y Condorcet y Laplace a problemas de aritmética política, existe durante este período escasa comunicación entre ambas disciplinas. Una contribución importante hacia dicha síntesis es debida a A. Quetelet (1846), que sostuvo la importancia del cálculo de probabilidades para el estudio de datos humanos. Quetelet demostró que la estatura de los reclutas de un reemplazo seguía una distribución normal, e introdujo el concepto de «hombre medio». Sin embargo, la diferencia de concepción y de lenguaje entre los matemáticos-astrónomos y los estadísticos-demógrafos dificultó la interacción entre ambos grupos. La unión entre ambas corrientes va a producirse a comienzos del siglo xx, favorecida, en gran parte, por los nuevos problemas teóricos y metodológicos que planteaba la contrastación empírica de la teoría de Darwin.

1.4.3 El nacimiento de la estadística actual

La revolución que supuso en la física Newton se produjo en la biología por la obra de Darwin. Dos facetas importantes de esta teoría eran: a) permitía establecer predicciones sobre la evolución de poblaciones animales que, en

determinadas condiciones, podían ser contrastadas empíricamente; b) la contrastación debería ser estadística, ya que la unidad que va a sufrir la evolución es la población en su conjunto. Los dos mecanismos de la selección natural, producción de variabilidad y selección mediante lucha por la existencia, tienen un atractivo inmediato desde el punto de vista estadístico. La producción de variabilidad mediante el azar entronca con el cálculo de probabilidades; la selección natural, con el estudio de poblaciones y con la idea de correlación. Aquellos organismos que estén más adaptados sobrevivirán un mayor período de tiempo y dejarán un mayor número de descendientes, por lo que tiene que existir una correlación entre determinadas características genéticas transmisibles y el grado de supervivencia y descendencia de los individuos de una especie.

El primero en resaltar la necesidad de acudir a métodos estadísticos para contrastar la teoría de Darwin fue Francis Galton (1822-1911). Galton, primo de Darwin, fue un hombre de profunda curiosidad intelectual que le llevó a viajar por todo el mundo y a realizar actividades tan diversas como redactar leyes para los hotentotes que gobernaban en el sur de África o realizar fecundas investigaciones en meteorología (a él le debemos el término «anticiclón»). La lectura de la obra de Darwin supuso una transformación radical en la vida de Galton, que, casi a los 40 años, dedica sus esfuerzos al estudio de la herencia humana. Su trabajo principal es *Natural Inheritance*, publicado en 1889 (a la edad de 67 años). Galton estudió exhaustivamente la distribución normal e introdujo el concepto de línea de regresión comparando las estaturas de padres e hijos. Galton encontró que los padres altos tenían, en promedio, hijos altos, pero en promedio más bajos que sus padres, mientras que los padres bajos tenían hijos bajos, pero, en promedio, más altos que sus padres. Este fenómeno de regresión se ha encontrado en muchas características hereditarias, de manera que los descendientes de personas extremas en alguna característica estarán, en promedio, más cerca de la media de la población que sus progenitores, produciendo así un efecto de regresión (vuelta) a la media de la población.

La importancia de Galton radica no solamente en el nuevo enfoque que introduce en el problema de la dependencia estadística, sino también en su influencia directa sobre Weldon, K. Pearson, R. A. Fisher y Edgeworth entre otros. El primer departamento de estadística en el sentido actual de la palabra fue patrocinado por él y llevó su nombre, y la revista *Biométrica* fue posible gracias a su generoso apoyo económico.

El enfoque estadístico propugnado por Galton para el estudio de los problemas de la evolución en *Natural Inheritance* es aceptado entusiásticamente por W. R. F. Weldon (1860-1906), entonces catedrático de zoología en la Universidad de Londres. Weldon abandona el camino de los estudios embriológicos y morfológicos como medio de contrastar las hipótesis de Darwin y comienza a investigar en la aplicación de los métodos estadísticos a la biología animal. En 1893 (Weldon, 1893), escribe:

Es necesario insistir en que el problema de la evolución animal es esencialmente un problema estadístico [...] debemos conocer: a) el porcentaje de animales que exhiben un cierto grado de anormalidad respecto a un carácter; b) el grado de anormalidad de otros órganos que acompaña a las normalidades de uno dado; c) la diferencia entre la tasa de mortalidad en animales con diferentes grados de anormalidad respecto a un órgano; d) la anormalidad de los descendientes en términos de anormalidad de los padres y viceversa.

La resolución de estos problemas requiere el desarrollo de métodos estadísticos más avanzados que los existentes, y Weldon busca para ello la colaboración de un matemático y filósofo: K. Pearson (1857-1936). La colaboración de estos dos autores y el apoyo de Galton van a constituir el impulso generador de la corriente de contribuciones que va a fundamentar la estadística actual.

El lector encontrará en los capítulos siguientes varias de las contribuciones de K. Pearson que llevan su nombre. Para facilitar la aplicación de los nuevos métodos, dados los escasos medios de cálculo disponibles a finales del siglo XIX, Pearson dedicó una parte importante de sus esfuerzos a la publicación de tablas estadísticas que permitieran la utilización práctica de los nuevos métodos, con lo que contribuyó, decisivamente, a su rápida difusión.

El laboratorio de K. Pearson se convierte en un polo de atracción para las personas interesadas en el análisis empírico de datos. W. S. Gosset (1876-1937), que trabajaba en la firma cervecera Guinness de Dublín, fue una de las personas que acudieron a Londres a estudiar bajo el patrocinio de Pearson. Gosset se había encontrado en sus investigaciones sobre los efectos de las características de la materia prima en la calidad de la cerveza final con el problema de las pequeñas muestras. No era posible económicamente, en este caso, obtener las grandes cantidades de datos que permitirían utilizar los métodos para muestras grandes desarrolladas por Pearson y su escuela. Para resolver el problema, Gosset realizó el primer trabajo de investigación estadística mediante el método de Montecarlo, tomando 750 muestras aleatorias de cuatro elementos de los datos recopilados por W. R. McDonnell sobre la estatura y la longitud del dedo corazón de 3.000 delincuentes, con los que simuló el proceso de tomar muestras de una distribución normal y obtuvo la distribución *t*, que publicó con el pseudónimo de Student, ya que Guinness no permitía divulgar las investigaciones de sus empleados.

Los fundamentos de la estadística actual y muchos de los métodos de inferencia expuestos en este libro son debidos a R. A. Fisher (1890-1962). Fisher se interesó primeramente por la eugenesia, lo que le conduce, siguiendo los pasos de Galton, a la investigación estadística. Sus trabajos culminan con la publicación de *Statistical Methods for Research Workers.* En él aparece ya claramente el cuerpo metodológico básico que constituye la

estadística actual: el problema de elegir un modelo a partir de datos empíricos, la deducción matemática de las propiedades del mismo (cálculo de probabilidades), la estimación de los parámetros condicionados a la bondad del modelo y la validación final del mismo mediante un contraste de hipótesis.

1.4.4 La expansión de la estadística durante el siglo xx

Entre 1920 y el final de la Segunda Guerra Mundial se extiende la aplicación de los métodos estadísticos en áreas tan diversas como la ingeniería (control de calidad por Shewart, métodos de predicción y control de procesos y codificación de señales por Wiener y Shannon), la economía (estimación de ecuaciones de oferta y demanda, índices de precios, medición de la riqueza y de la pobreza), la física (teoría cinética de los gases), la antropología (clasificación de restos arqueológicos), la psicología (medición de la inteligencia y teoría de test) o la medicina (pruebas para determinar la eficacia de nuevos tratamientos). La búsqueda de respuestas a los nuevos interrogantes planteados por estas aplicaciones impulsan, a su vez, el desarrollo de nuevos métodos estadísticos. Los problemas en agronomía conducen a Fisher a crear la teoría de diseños experimentales, y un problema de discriminación en antropología (concretamente la clasificación de cráneos; véase J. Box, 1978) lleva a Fisher a inventar el análisis discriminante. El análisis factorial surge ligado a problemas en la psicología, y, en general, la economía y las ciencias sociales impulsan el desarrollo de métodos para medir la relación entre variables (métodos de regresión) y analizar muchas variables conjuntamente (métodos multivariantes). Los problemas de ingeniería conducen a un estudio sistemático de la teoría de modelos dinámicos (procesos estocásticos) y a la creación de la teoría de predicción y de extracción de señales de Wiener y Kolmogorov en los años cuarenta. Las necesidades en el control de procesos sugieren a E. S. Pearson la creación de la teoría general de contraste de hipótesis, conjuntamente con J. Neyman, y el trabajo en aplicaciones industriales de la estadística del Statistical Research Group en Columbia durante la Segunda Guerra Mundial condujo a Wald a inventar los contrastes secuenciales para el control de recepción, punto de partida básico en el nacimiento y desarrollo de la teoría estadística de decisión.

A partir de 1950 podemos considerar que comienza la época moderna de la estadística. Algunos de los aspectos diferenciales respecto a los períodos anteriores son:

a) La aparición del ordenador digital, que va a revolucionar la metodología estadística y abrir enormes posibilidades para la construcción de modelos más complejos.

b) El cambio de énfasis en la metodología estadística. La influencia de Neyman, Pearson y Wald en los años cuarenta y cincuenta concentra la investigación teórica en la búsqueda de procedimientos óptimos de estimación y contraste de hipótesis en problemas simplificados. Estos procedimientos óptimos parten de dos premisas fundamentales: a) Los datos de que disponemos han sido generados por un modelo de distribución de probabilidad que es conocido salvo por un vector de parámetros; b) el modelo pertenece a una familia restringida de distribuciones de probabilidad, matemáticamente tratable. Este análisis considera tangencialmente dos problemas centrales del análisis estadístico: a) la identificación de la estructura del modelo y b) los contrastes diagnósticos para, una vez estimado el modelo, decidir si puede rechazarse su estructura básica mediante los datos empíricos.
El enfoque de «métodos óptimos» se apoya en el postulado de continuidad: una pequeña desviación en las hipótesis producirá una pequeña variación en el resultado final. La falsedad de este principio en muchos problemas estadísticos relevantes ha conducido a que actualmente la metodología estadística ponga el énfasis en el proceso iterativo de aprendizaje a partir de los datos en lugar de en la aplicación de un determinado procedimiento óptimo. Con esto, las fases de exploración de los datos, de consideración de modelos flexibles y de procedimientos robustos y generales de estimación han pasado a ocupar el centro de la metodología moderna.

A finales del siglo XX el espectacular aumento de la capacidad de cálculo de los ordenadores y la caída de los costes de almacenamiento de la información han hecho posible la recogida automática de grandes masas de datos en cualquier actividad humana. Por ejemplo, los datos que los satélites nos envían en un solo día bloquearían la capacidad de análisis de un ordenador de los años setenta; un banco, un supermercado o una tienda virtual en Internet adquiere, a través de las operaciones por tarjeta de crédito de sus clientes, bancos de datos del comportamiento de los consumidores que serían intratables hace pocos años, y los procesos de fabricación automática proporcionan información constante de muchas variables de control que, además, evolucionan dinámicamente en el tiempo. El reto más importante de la estadística en este siglo XXI es cómo extraer la información en estas grandes masas de datos y utilizarla de manera efectiva para aumentar nuestro conocimiento, orientar la toma de decisiones y dirigir la mejora de procesos y servicios.

Esta breve revisión de la evolución de la estadística muestra cómo los grandes períodos de avances teóricos se han producido generalmente ligados a la resolución de importantes problemas prácticos. Paradójicamente, la mayoría de las contribuciones importantes que hemos revisado son debidas

a investigadores que no pueden calificarse exclusivamente como estadísticos, sino, en el sentido más amplio del término, como científicos. Además, los avances más importantes en algunas disciplinas se han producido por la utilización de métodos generados para resolver los problemas en otra. Por ejemplo, métodos estadísticos desarrollados para estudiar los procesos de difusión en física están revolucionando la investigación financiera (sus contribuciones han sido reconocidas ya con un premio Nobel); herramientas desarrolladas en ingeniería espacial (el filtro de Kalman) son ya de uso común en el control automático de procesos industriales, en la predicción de magnitudes macroeconómicas y en el tratamiento digitalizado de imágenes tomadas por un escáner en medicina; y métodos de clasificación desarrollados en antropología son la base de los sistemas de concesión automática de créditos, de las máquinas que reconocen billetes y monedas, de los sistemas de reconocimiento de voz y de la construcción de buscadores eficaces en Internet.

1.5 Lecturas recomendadas

Un libro excelente sobre cómo prevenir la manipulación con la estadística es Huff (1993). Bartholomew y Bassett (1971) y Moroney (1990) presentan introducciones no técnicas a la utilización de la estadística en el mundo de hoy. Tanur *et al.* (2007) es una colección muy interesante de aplicaciones. Borel (1998) es un breve y delicioso ensayo sobre las probabilidades en la vida diaria. Huff (1993) y Kitaigorodski (1976) presentan numerosos ejemplos en la misma dirección. Rao (2004) es un interesante ensayo sobre el papel de la estadística en la investigación científica.

El lector interesado en la historia del cálculo de probabilidades puede acudir a David (1998) y Todhunter (2007). La historia de la estadística se presenta en Pearson y Kendall (1976), Kendall y Plackett (1977), Stigler (1990, 2002) y Hald (1990, 1998). Sánchez-Lafuente (1975) estudia la historia de la estadística en España hasta 1900. Biografías de estadísticos célebres de especial interés son Box (1978), que describe la vida de Fisher, Reid (1982), que narra la de Neyman, o Pearson (1990), que se centra en la de Student, y Sánchez y Valdés (2003) la de Kolmogorov. Gani (1982) contiene ensayos autobiográficos de estadísticos actuales. El libro de Bernstein (1998) es una historia fascinante del riesgo y la incertidumbre.

Primera parte

Datos

2. La descripción de una variable

Pierre-Simon, marqués de Laplace (1749-1827)
Científico francés y uno de los creadores de la teoría de la probabilidad. Hizo también contribuciones fundamentales a la física, la astronomía y las matemáticas. Fue profesor de Napoleón en la escuela militar, y aquél le eligió como ministro del Interior. Luis XVIII le hizo marqués en la restauración monárquica.

2.1 Datos y distribuciones de frecuencias

Dado un conjunto de datos de una variable x, la estadística descriptiva estudia procedimientos para sintetizar la información que contienen.

Los tipos de variables que consideraremos son:

a) *Variables cualitativas, categóricas o atributos:* no toman valores numéricos y describen cualidades. Por ejemplo, clasificar personas por el color de su pelo.
b) *Variables cuantitativas discretas:* toman únicamente valores enteros; corresponden en general a contar el número de veces que ocurre un suceso. Por ejemplo, número de compras de un producto en un mes.
c) *Variables cuantitativas continuas:* toman valores en un intervalo; corresponden a medir magnitudes continuas. Por ejemplo, tiempo entre la llegada de dos autobuses.

Supondremos que el orden en que se recogen los datos es irrelevante. Cuando los datos se observan con una pauta temporal fija (cada mes, año, etc.), constituyen una serie temporal y su análisis requiere métodos especiales que tengan en cuenta que el orden de los datos es informativo.

La presentación de un conjunto de datos suele hacerse indicando los valores de la variable y sus frecuencias de aparición, tanto absolutas como relativas. La *frecuencia relativa* de un suceso A se define por:

$$fr(A) = \frac{\text{número de veces que se observa } A}{\text{número total de datos}}$$

2.1.1 Distribuciones de frecuencias

La tabla 2.1 presenta un ejemplo de una distribución de frecuencias para una variable cualitativa: se indican las clases o atributos y sus frecuencias observadas. Cuando los atributos no corresponden a una escala ordinal (por ejemplo alto, medio, bajo), conviene ordenarlos por su frecuencia de aparición.

Tabla 2.1 Distribución de defectos en libros en una imprenta

Clases	Frecuencia	Frecuencia relativa
Corte de las hojas	60	0,43
Mala impresión	40	0,29
Tinta irregular	20	0,14
Encuadernación	12	0,09
Portada	6	0,04
Lomo	2	0,01
TOTAL	140	1

La tabla 2.2 presenta esta misma idea para una variable discreta. Esta representación es útil cuando el número de valores posibles es pequeño. En otro caso, conviene agrupar los datos, como se aprecia en la mencionada tabla.

Agrupamiento

Cuando el número de valores distintos que toma una variable discreta sea grande, o cuando ésta sea continua, conviene agrupar los datos en clases, como sigue:

Tabla 2.2 Distribución de frecuencias de la variable: número de llamadas recibidas en una centralita en períodos de un minuto

X	(*f*) frecuencia	(*fr*) frecuencia relativa
0	40	0,44
1	26	0,29
2	14	0,16
3	6	0,07
4	3	0,03
5	0	0,00
6	1	0,01
TOTAL	90	1

a) Redondear los datos a dos o, a lo sumo, tres cifras significativas eligiendo las unidades para que cada observación contenga dos o tres dígitos, sin coma decimal.
b) Decidir el número *r* de clases a considerar. Este número debe ser entre 5 y 20. Una regla frecuentemente utilizada es tomar *r* igual al entero más próximo a $\sqrt{n}$, siendo *n* el número de datos. Esta regla es indicativa y conviene probar con distinto número de clases y escoger aquel que proporcione una descripción más clara.
c) Seleccionar los límites de clase que definen los intervalos, de manera que las clases sean de la misma longitud y cada observación se clasifique sin ambigüedad en una sola clase.
d) Contar el número de observaciones en cada clase, que llamaremos la frecuencia de clase, y obtener la frecuencia relativa de cada clase dividiendo aquélla por el total de datos.

La tabla 2.3 presenta un ejemplo de una distribución de frecuencias para una variable continua. Llamaremos en adelante *marca de clase* al centro del intervalo que define la clase.

2.1.2 Diagramas de tallo y hojas

Un procedimiento semigráfico de presentar la información para variables cuantitativas, que es especialmente útil cuando el número total de datos es pequeño (menor que 50), es el diagrama de tallo y hojas de Tukey. Los principios para construirlo son:

Tabla 2.3 Distribución de la variable: tiempo en minutos al realizar una operación

Intervalo	Centro del intervalo	Frecuencia relativa
20-24	22	0,30
25-29	27	0,40
30-34	32	0,20
35-39	37	0,07
40-44	42	0,03

a) Redondear los datos a dos o tres cifras significativas, expresándolos en unidades convenientes.
b) Disponerlos en una tabla con dos columnas separadas por una línea como sigue:

   b.1) Para datos con dos dígitos, escribir a la izquierda de la línea los dígitos de las decenas —que forma el tallo— y a la derecha las unidades, que serán las hojas. Por ejemplo, 87 se escribe 8|7.
   b.2) Para datos con tres dígitos el tallo estará formado por los dígitos de las centenas y decenas, que se escribirán a la izquierda, separados de las unidades. Por ejemplo, 127 será 12|7.

c) Cada tallo define una clase, y se escribe sólo una vez. El número de «hojas» representa la frecuencia de dicha clase.

La tabla 2.4 presenta un ejemplo de estos diagramas.

Cuando el primer dígito de la clasificación varía poco, la mayoría de los datos tienden a agruparse alrededor de un tallo y el diagrama resultante tiene poco detalle. En ese caso es conveniente subdividir cada tallo en dos o más partes introduciendo algún signo arbitrario, como se indica en la tabla 2.5.

2.2 Representaciones gráficas

2.2.1 Diagrama de Pareto

Este diagrama se utiliza para representar datos cualitativos y se construye como sigue:

Tabla 2.4 Diagrama de tallo y hojas

(1) Datos recogidos en cm:
11,357; 12,542; 11,384; 12,431; 14,212; 15,213; 13,300; 11,300; 17,206; 12,710; 13,455; 16,143; 12,162; 12,721; 13,420; 14,698.

(2) Datos redondeados expresados en mm:
114; 125; 114; 124; 142; 152; 133; 113; 172; 127; 135; 161; 122; 127; 134; 147.

(3) Diagrama de tallo y hojas, datos en mm:

11	443
12	54727
13	354
14	27
15	2
16	1
17	2
decenas	unidades

1) Se ordenan las categorías o clases por su frecuencia relativa de aparición.
2) Cada categoría se representa por un rectángulo cuya altura es su frecuencia relativa.

La figura 2.1 presenta el diagrama de Pareto para los tipos de defectos encontrados en libros de la tabla 2.1. Se observa que la mayoría de los defectos (casi tres cuartas partes) corresponden a unas pocas clases (casi una cuarta parte). Este resultado se conoce como «ley de Pareto» y se observa aproximadamente en muchos campos tan distintos como la economía (distribución de la riqueza, de los beneficios empresariales), la geografía (tamaño de ríos, montañas, ciudades), la ingeniería (tipos de defectos, averías, etc.) o la lingüística (frecuencia de uso de las palabras en un idioma).

2.2.2 Diagrama de barras

Para datos de variables discretas, y en general para distribuciones de frecuencias de datos sin agrupar, se utiliza el *diagrama de barras.* Este diagrama representa los valores de la variable en el eje de abcisas levantando en cada punto una barra de longitud igual a la frecuencia relativa. La figura 2.2 representa el diagrama de barras asociado a la tabla 2.2.

Figura 2.1 Diagrama de Pareto para la tabla 2.1

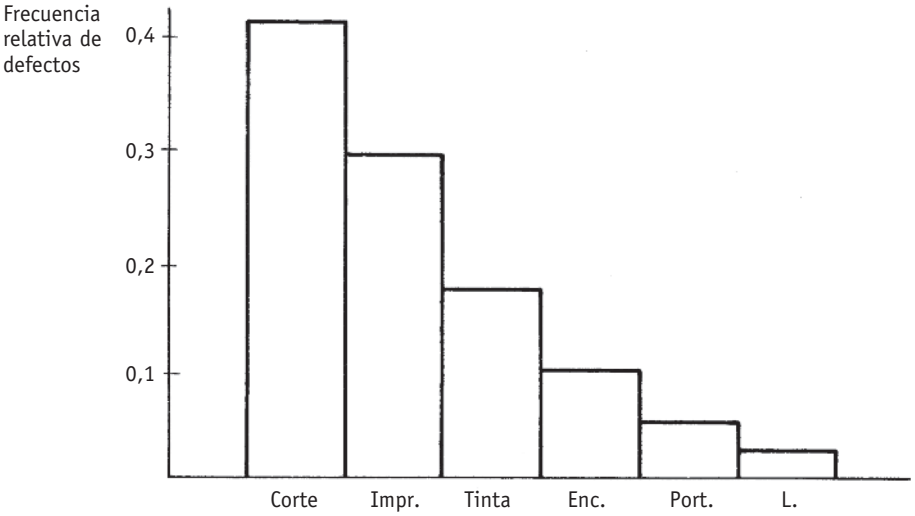


Figura 2.2 Diagrama de barras de la tabla 2.2

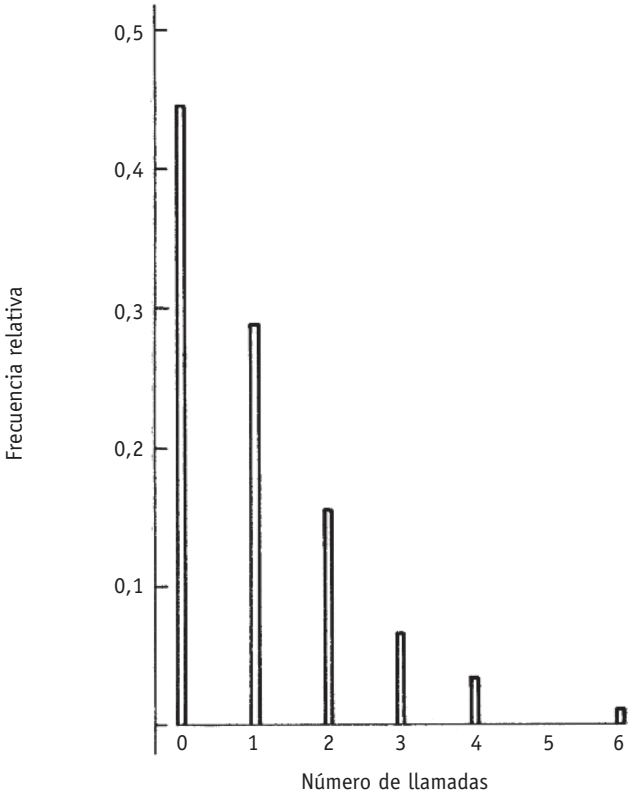


Tabla 2.5 Diagrama de tallo y hojas con subdivisión del tallo

(1) Las pulsaciones por minuto de un grupo de 40 personas se han representado en el diagrama de tallo y hojas siguiente:

5	2 6
6	0 0 0 0 0 0 4 4 4 4 4 4 8 8 8 8 8 8 8 8
7	2 2 2 2 2 2 2 2 6 6 6 6 6 6
8	0 0 4 4 8 8
9	2

(2) Podemos obtener más detalle subdividiendo cada tallo en dos partes iguales: en una colocaremos las hojas 0 a 4 y lo representamos por (∗), y en la otra, las hojas de 5 a 9 y lo representaremos por (·), obteniendo el diagrama:

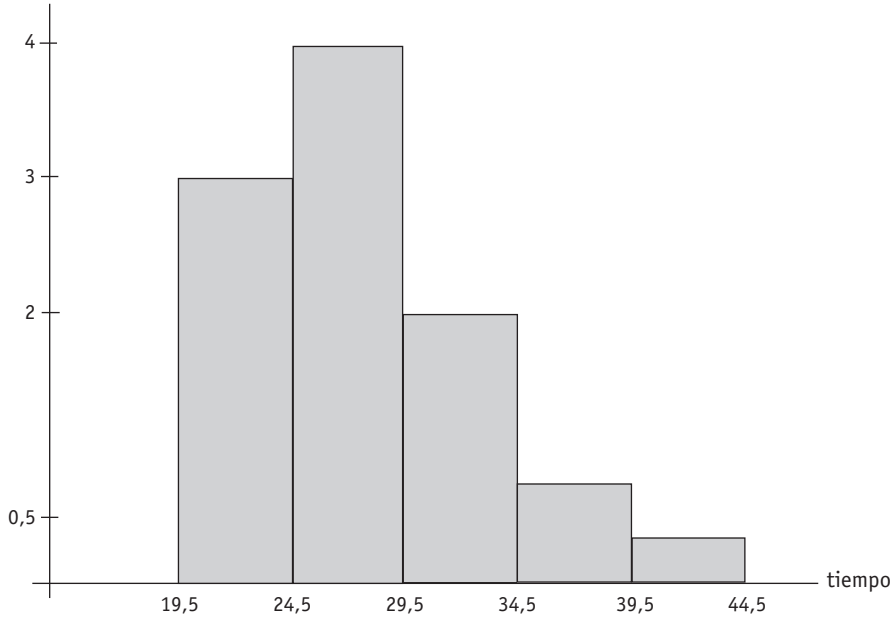
5 ∗	2
·	6
6 ∗	0 0 0 0 0 0 4 4 4 4 4 4
·	8 8 8 8 8 8 8 8
7 ∗	2 2 2 2 2 2 2 2
·	6 6 6 6 6
8 ∗	0 0 4 4
·	8 8
9 ∗	2

Observemos que todos los datos son múltiplos de 4, lo que hace sospechar que se han obtenido midiendo las pulsaciones cada 15 segundos y multiplicando por cuatro.

2.2.3 Histogramas

La representación gráfica más frecuente para datos agrupados es el histograma. Un *histograma* es un conjunto de rectángulos, cada uno de los cuales representa un intervalo de agrupación o clase. Sus bases son iguales a la amplitud del intervalo, y las alturas se determinan de manera que su área sea proporcional a la frecuencia de cada clase. La figura 2.3 representa el histograma asociado a los datos agrupados de la tabla 2.3; a efectos de representación se considera que el intervalo (20-24) comprende valores desde (19,5 a 24,5), el siguiente desde (24,5-29,5), etc. De esta manera se abarca todo el campo de la variación de la variable sin dejar huecos y, al mismo tiempo, cada observación se clasifica en sólo una clase sin ambigüedad. (Estamos suponiendo que los datos originales son números enteros.)

Figura 2.3 Histograma de los datos de la tabla 2.3



Los histogramas pueden proporcionar mucha información respecto a la estructura de los datos, y la figura 2.4 presenta varios casos típicos.

El histograma (a) de la figura 2.4 presenta una distribución asimétrica que es típica de datos económicos, y en general de mediciones de distribuciones de renta, población, consumo de electricidad, tamaño de empresas, etc.; el histograma (b) muestra una distribución simétrica que aparece en muchos procesos de fabricación al estudiar la distribución de una medida de calidad; el histograma (c) aparece al mezclar elementos de varias poblaciones cada uno de ellos con distribución tipo (b), lo que produce una distribución con gran variabilidad. En el límite, si las distribuciones individuales están muy separadas, podemos encontrarnos una situación como la descrita por el histograma (d), donde se apuntan más claramente ambas distribuciones. El caso (e) representa una distribución truncada, que aparecerá, por ejemplo, al medir el peso de ciertos elementos en un control de calidad que tiene límites de especificaciones *A* y *B*. Finalmente, la distribución (f) es muy asimétrica y surge al estudiar tiempos entre averías, entre llegadas, entre accidentes, etc.

Cuando el número de datos es pequeño, una representación más útil que el histograma es el *diagrama de puntos.* Por ejemplo, la figura 2.5 presenta gráficamente los datos del diagrama de hojas y tallos de la tabla 2.4.

Figura 2.4 Algunos histogramas típicos

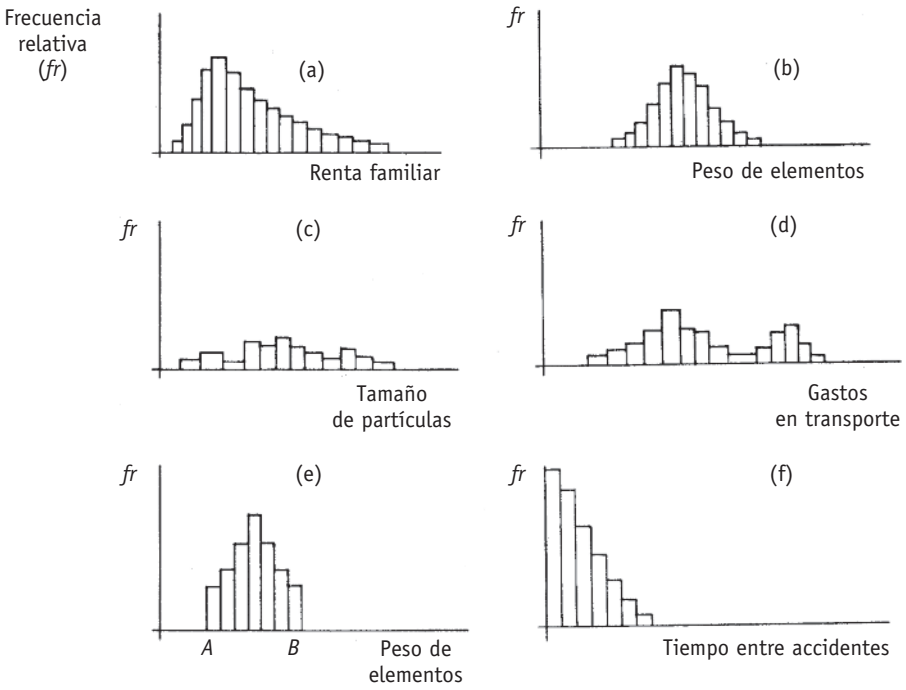
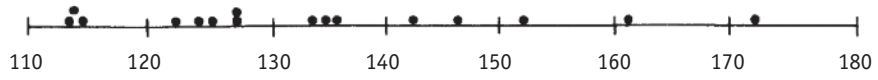


Figura 2.5 Diagrama de puntos para los datos de la tabla 2.4



2.2.4 Gráficos temporales

Cuando se observa una variable a intervalos regulares de tiempo (día, mes, año, etc.), la secuencia de valores constituye una *serie temporal.* Las figuras 2.6 y 2.7 presentan ejemplos de series temporales. Se observa que los datos próximos en el tiempo se parecen entre sí más que los muy alejados. Este fenómeno, característico de las series temporales, hace que el orden de los datos sea importante y deba tenerse en cuenta en el análisis.

Los procedimientos de análisis de datos que estudiaremos en este primer tomo *se aplican a datos cuya secuencia es irrelevante,* y no son válidos por

Figura 2.6 Consumo de energía eléctrica en España (1963-1988)

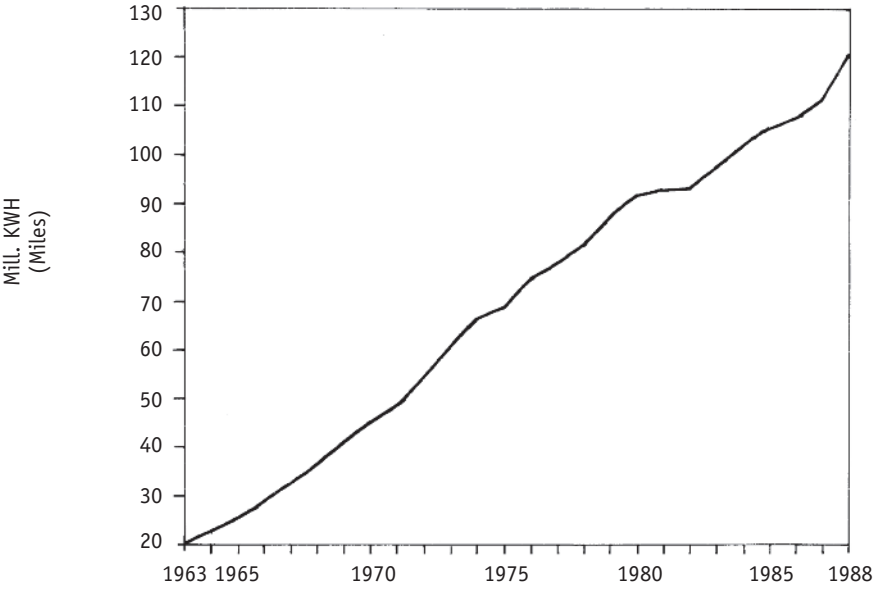
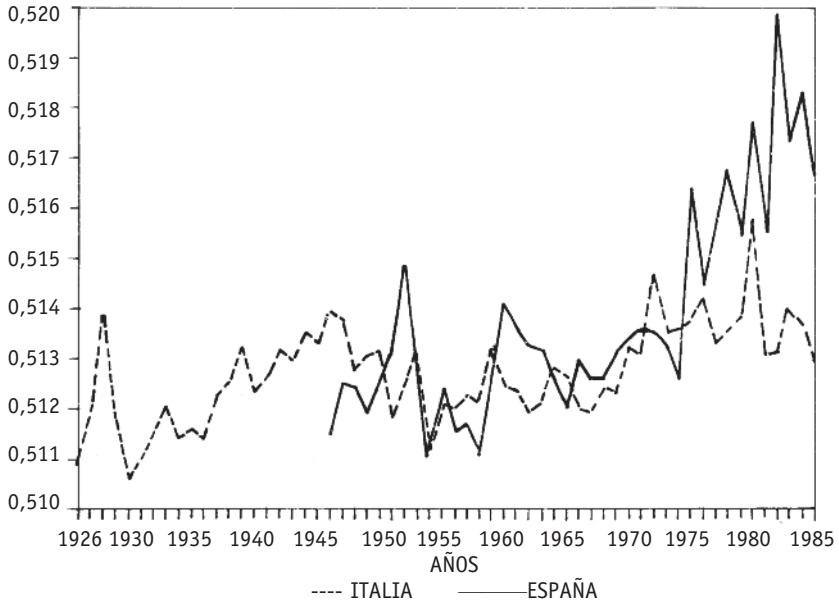


Figura 2.7 Proporción de nacidos varones sobre el total de nacimientos en Italia (línea de trazos) y España (línea continua)



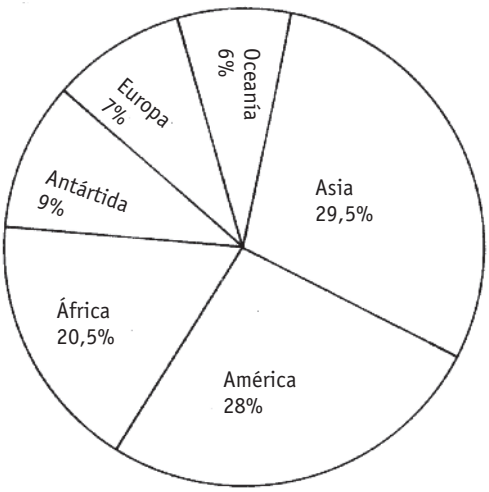
tanto para series temporales. En el apéndice 2B se presenta una breve introducción al estudio descriptivo de series temporales.

2.2.5 Otras representaciones gráficas

El objetivo de un gráfico es describir simple y fielmente la información contenida en los datos observados. En consecuencia, la naturaleza de la variable estudiada puede sugerir una representación gráfica específica distinta de las anteriores.

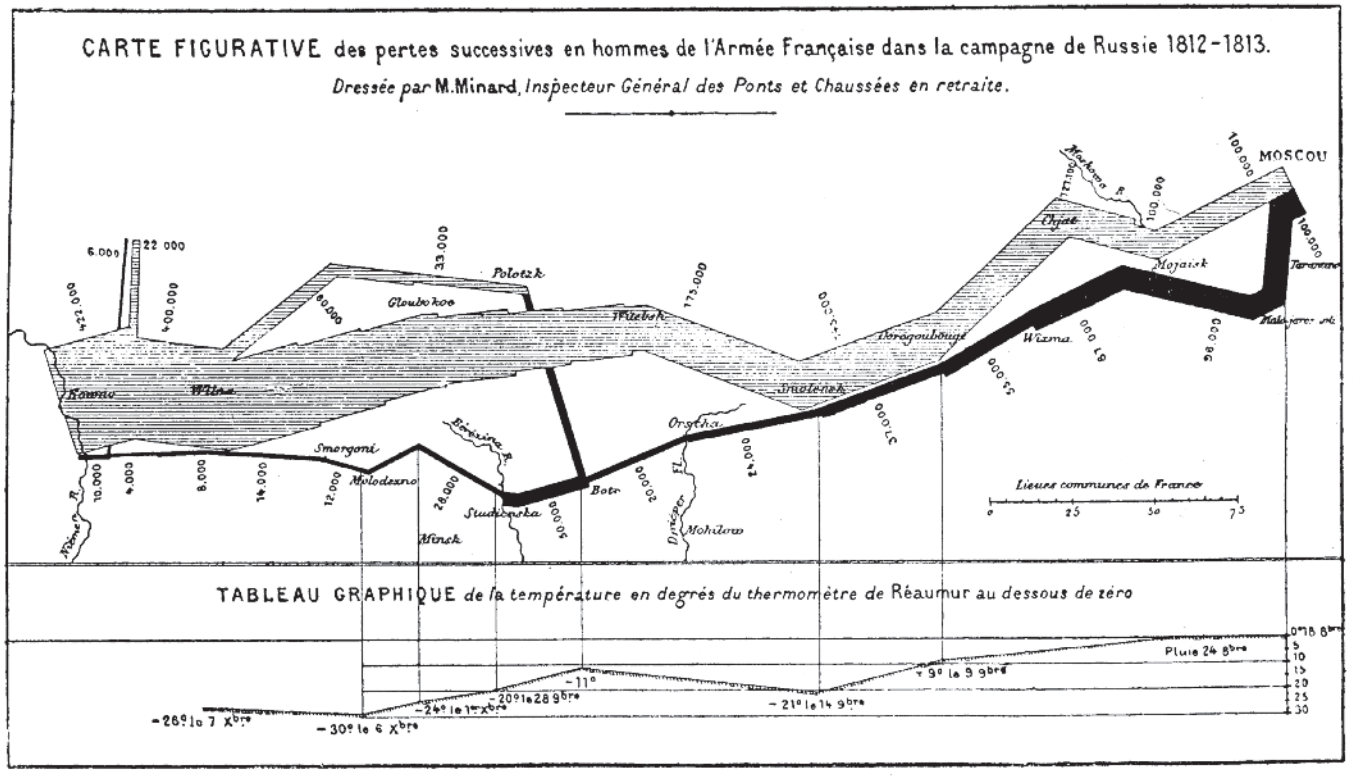
Por ejemplo, para reflejar la idea de división de un conjunto en categorías excluyentes se utilizan como alternativa a los gráficos de Pareto los diagramas de tarta, que se construyen de manera que el área de cada porción sea proporcional a la frecuencia relativa (véase la figura 2.8). Estas representaciones se denominan *pictogramas.*

Figura 2.8 Proporción de superficie ocupada por los distintos continentes



La figura 2.9 representa simultáneamente tres series temporales utilizando conjuntamente la dimensión espacial y temporal. El gráfico escogido permite ilustrar el tamaño del ejército (por la anchura del trazo) de Napoleón en la campaña de Rusia a lo largo del tiempo y su posición, conjuntamente con la temperatura durante su retirada en el invierno de 1812. Este gráfico es un buen ejemplo de cómo representar los datos relevantes de una realidad compleja.

Figura 2.9 **La campaña de Rusia de Napoleón. Tomado de E. J. Marey, *La Méthode Graphique* (París, 1885). Reproducido con autorización de Tufte (1983)**



2.3 Medidas de centralización y dispersión

Cuando disponemos de un conjunto de datos homogéneo (su orden es irrelevante) de una variable cuantitativa, resulta conveniente complementar la distribución de frecuencias con ciertas medidas resumen. Las más importantes son las de *tendencia central* o *centralización,* que indican el valor medio de los datos, y las de *dispersión,* que miden su variabilidad. En la sección siguiente estudiaremos medidas que describen la forma de la distribución, como su *grado de simetría o de concentración* de la distribución.

Es importante tener en cuenta que las medidas resumen son informativas para datos homogéneos y que pueden ser muy engañosas cuando mezclamos distintas poblaciones. Por ejemplo, si el histograma de los datos es del tipo 2.4(d), una medida «media» del valor de los datos no representará a ninguna de las dos subpoblaciones. En estos casos es más adecuado identificar las razones de la heterogeneidad, dividir los datos en dos poblaciones distintas y calcular las medidas características en cada una de ellas.

2.3.1 Medidas de centralización

Media

Dado un conjunto de datos numéricos $x_1, ..., x_n$, se define la media aritmética por:

$$\bar{x} = \frac{x_1 + ... + x_n}{n} = \frac{\Sigma x_i}{n} \qquad (2.1)$$

donde el símbolo Σ, que se denomina sumatorio, quiere decir que debemos sumar todos los valores de la variable. Para datos discretos agrupados, como en la tabla 2.2, llamando x_j a los valores distintos de la variable y $fr(x_j)$ a sus frecuencias relativas respectivas, el cálculo de la media se efectúa con:

$$\bar{x} = \sum x_j fr\,(x_j) \qquad (2.2)$$

donde el sumatorio va extendido ahora al número de valores distintos de la variable.

Para datos agrupados en clases, la media se calcula suponiendo que todos los datos de cada clase son idénticos al centro de la clase, con lo que, llamando m_j a estos valores centrales y $fr(m_j)$ a la frecuencia relativa de la clase j, la fórmula se reduce a:

$$\bar{x} = \sum m_j fr\,(mj) \qquad (2.3)$$

donde, como en (2.2), el sumatorio va extendido al número total de clases.

La media aritmética es el centro de los datos en el sentido de equilibrar los valores por defecto y por exceso respecto a la media. En otros términos, la suma de las desviaciones de los datos con relación a la media toma el valor mínimo, cero. Para comprobarlo supongamos datos sin agrupar. Entonces:

$$\sum (x_i - \bar{x}) = \sum x_i - n\bar{x} = 0$$

La media es, en este sentido, el centro geométrico o «centro de gravedad» del conjunto de datos de la variable. Además, la suma de las desviaciones al cuadrado entre los datos y la media es mínima, es decir, la media es el valor *a* que minimiza

$$\sum (x_i - a)^2$$

En efecto, derivando respecto a *a* se obtiene la condición $\Sigma\,(x_i - a) = 0$, que implica que *a* debe ser la media aritmética.

Mediana y moda

La *mediana* es un valor tal que, ordenados en magnitud los datos, el 50% es menor que ella y el 50% mayor. Por tanto, al ordenar los datos sin agrupar, la mediana es el valor central, si su número es impar, o la media de los dos centrales, si hay un número par. Para datos agrupados discretos se toma como mediana el valor x_m tal que $fr(x < x_m) < 0{,}5$ pero $fr(x \leq x_m) \geq 0{,}5$. Es decir, si ordenamos los valores de la variable antes de x_m tenemos menos del 50% de los datos, pero al incluir x_m tenemos al menos el 50%. Para datos continuos agrupados en intervalos se toma como mediana el centro del «intervalo central» (x_a, x_b) que verifica:

$$\begin{aligned} fr\,(x \leq x_a) &< 0{,}5 \\ fr\,(x \leq x_b) &> 0{,}5 \end{aligned}$$

La *moda* es el valor más frecuente.

El uso de las medidas de centralización

Conviene calcular las medidas de centralización sobre datos homogéneos, ya que comparar los valores de poblaciones heterogéneas puede ser muy

engañoso. Por ejemplo, supongamos que se calcula el tiempo medio que un estudiante requiere para completar una carrera universitaria en dos universidades U_1 y U_2 obteniendo 5,5 años en ambas. ¿Podemos concluir que son igualmente difíciles? Es posible que no. Supongamos que la universidad U_1 es muy homogénea y contiene sólo facultades con títulos de cinco años y dificultad análoga que los alumnos completan en 5,5 años de promedio. Sin embargo, la U_2 es más heterogénea, con carreras de cuatro años, que requieren en promedio 5 años, y carreras de seis, que requieren en promedio 7,5 años. Supongamos que en U_2 el 80% de los estudiantes cursan las primeras y el 20% las segundas, con lo que resulta la duración media de:

$$\bar{x} = 0{,}80\ (5) + 0{,}2\ (7{,}5) = 5{,}5 \text{ años}$$

En la primera universidad los estudiantes invierten medio año más de lo previsto, mientras que en la segunda invierten entre un año y año y medio más de lo previsto, con lo que es razonable admitir una mayor dificultad en la segunda. Este problema aparece por comparar medias de situaciones heterogéneas.

Aunque desde un punto de vista puramente descriptivo las tres medidas de centralización estudiadas proporcionan información complementaria, sus propiedades son muy distintas: la media utiliza todos los datos y es, por tanto, preferible si los datos son homogéneos; tiene el inconveniente de que es muy sensible a observaciones atípicas, y un error de datos o un valor anormal puede modificarla totalmente. Por el contrario, la mediana utiliza menos información que la media, ya que sólo tiene en cuenta el orden de los datos y no su magnitud, pero, en contrapartida, no se ve alterada si una observación —o en general una pequeña parte de las observaciones— contiene errores grandes de medida o de transcripción.

En consecuencia, es siempre recomendable calcular la media y la mediana: ambas medidas diferirán mucho cuando la distribución sea muy asimétrica, lo que sugiere heterogeneidad en los datos.

Ejemplo 2.1

Las medidas de centralización para la distribución de frecuencias de la tabla 2.2 son:

media = $\bar{x}$ = 0.0,44 + 1 · 0,29 + 2.0,16 + 3.0,07 + 4.0,03 + 6.0,01 = 1

mediana = 1, ya que $fr\ (x < 1) = fr\ (x = 0) = 0,44 < 0,5$, mientras que $fr\ (x \leq 1) = 0,73 > 0,5$

moda = 0

y para la distribución de la tabla 2.3:

media = x = 22.0,3 + 27.0,4 + 32.0,2 + 37.0,07 + 42.0,03 = 27,65
mediana = 27, ya que $fr\ (x \leq 25) = 0{,}3 < 0{,}5$, mientras que
$fr\ (x \leq 29) = 0{,}7 > 0{,}5$
moda = 27

2.3.2 Medidas de dispersión

Desviación típica

A cada medida de centralización podemos asociarle una medida de la variabilidad de los datos respecto a ella. Para ello calculamos un promedio de las desviaciones de los datos respecto a la medida de centralización. A la media le asociamos la desviación típica, o desviación estandar. Para obtenerla se calcula la desviación de cada dato respecto a su media, se elevan al cuadrado estas desviaciones para que sean positivas y se promedian. A continuación se extrae la raíz cuadrada y el resultado es la desviación típica. Para datos sin agrupar se calcula por:

$$s = \sqrt{\frac{\Sigma\,(x_i - \bar{x})^2}{n}} \tag{2.4}$$

La desviación típica es un promedio de las desviaciones de los datos respecto a su media. Las desviaciones $(x_i - \bar{x})$ se elevan al cuadrado para convertirlas en positivas [recuérdese que $E(x_i - \bar{x}) = 0$] y se extrae la raíz cuadrada de su promedio para que la medida resultante tenga las mismas dimensiones que los datos originales. Su cuadrado se denomina *varianza*.

Para datos agrupados, la fórmula para la desviación típica se reduce a:

$$s = \sqrt{\sum\,(x_j - \bar{x})^2\, fr\,(x_j)} \tag{2.5}$$

donde el sumario se extiende ahora al número de clases o número de valores *distintos* de la variable. Para datos agrupados en intervalos, la fórmula es idéntica a la (2.5) sustituyendo x_j por el centro del intervalo, m_j.

Ejemplo 2.2

Calcular las desviaciones típicas para los datos de las tablas 2.2 y 2.3.

Comenzando con los datos de la tabla 2.2, como la media es 1 (véase ejercicio 2.1), la varianza será:

$$s^2 = (0 - 1)^2\, 0{,}44 + (2 - 1)^2\, 0{,}16 + (3 - 1)^2\, 0{,}07 + (4 - 1)^2\, 0{,}03 + \\ + (6 - 1)^2\, 0{,}01 = 1{,}4$$

$$s = \sqrt{1{,}4} = 1{,}18$$

La desviación típica es 1,18. El lector debe volver a los datos de la tabla 2.2 y comprobar que este valor representa la desviación promedio respecto a la media de 1. Por ejemplo, en esta tabla los valores que se alejan más de una desviación típica de la media son 3, 4, 5 y 6, que tienen en conjunto una probabilidad pequeña, de 0,11. Otra manera de verlo es darse cuenta de que la desviación a la media es cero el 29% de las veces (cuando $x = 1$), uno el 60% de las veces (cuando $x = 0$ y $x = 2$) y mayor de uno (2, 3, 4 o 5) el 11% de las veces.

Para los datos de la tabla 2.3, como la media es 27,65 (ejemplo 2.1), calcularemos primero la varianza:

$$s^2 = \Sigma\, (x_j - x)^2\, fr(x_j) = (22 - 27{,}65)^2\, 0{,}3 + (27 - 27{,}65)^2\, 0{,}4 + \\ + (32 - 27{,}65)^2\, 0{,}2 + (37 - 27{,}65)^2\, 0{,}07 + (42 - 27{,}65)^2\, 0{,}03 = 25{,}83$$

$$s = \sqrt{25{,}83} = 5{,}08$$

La desviación típica son 5,08 minutos. En la tabla observamos que más del 50% de las veces el viaje se desvía menos de una desviación típica de la media.

Interpretación de la desviación típica

La información conjunta que proporcionan la media y la desviación típica puede precisarse de la siguiente forma: entre la media y k veces la desviación típica existe, como mínimo, el

$$100\left(1 - \frac{1}{k^2}\right)\%$$

de las observaciones. Por ejemplo, si la media es 500 y la desviación típica 20, entre la media y 2 desviaciones típicas, es decir, entre 460 y 540 estarán, como mínimo, el

$$100\left(1-\frac{1}{4}\right)\% = 75\%$$

de las observaciones, y entre la media y 3 desviaciones típicas —entre 440 y 560— estarán como mínimo el

$$100\left(1-\frac{1}{9}\right)\% = 89\%$$

La demostración de esta propiedad es inmediata; partiendo de la definición de *s* dividamos los datos en dos clases: en la primera pondremos aquellas observaciones situadas a una distancia de la media mayor que *ks,* y que diremos pertenecen a la clase A_1; en la segunda estarán el resto de las observaciones, que no verifican esa propiedad, y que pertenecerán a la clase complementaria A_2. Entonces:

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n} = \sum_{A_1} \frac{(x_i - \overline{x})^2}{n} + \sum_{A_2} \frac{(x_i - \overline{x})^2}{n} \geq \sum_{A_1} \frac{(x_i - \overline{x})^2}{n}$$

ya que el segundo sumando es siempre positivo. Sustituyendo ahora, en cada término del conjunto $A_1 (x_i - \overline{x})^2$ por $k^2 s^2$ que, por construcción, es menor que cada uno de ellos, el sumatorio al tener todos los términos iguales será igual al valor común por el número de términos del sumatorio. El número de términos del conjunto A_1 es, por construcción, el número de observaciones con distancia a la media mayor que *ks* y:

$$s^2 \geq \sum_{A_1} \frac{(x_i - \overline{x})^2}{n} > k^2 s^2 fr\,(|x_i - \overline{x}| > ks);$$

por tanto, concluimos que:

$$fr(|x_i - \overline{x}| > ks) < \frac{1}{k^2}$$

que equivale a:

$$\boxed{fr(|x_i - \overline{x}| \geq ks) \geq 1 - \frac{1}{k^2}} \qquad (2.6)$$

que nos permite concluir que, en cualquier distribución, se encuentran, al menos:

Entre la media y dos desviaciones típicas el 75%
Entre la media y tres desviaciones típicas el 89%

Esta desigualdad se denomina *desigualdad de Tchebychev.*

Coeficiente de variación

Se denomina coeficiente de variación al cociente

$$CV = \frac{s}{|\bar{x}|} \tag{2.7}$$

donde suponemos que $\bar{x} \neq 0$ y $|\bar{x}|$ es el valor absoluto de $\bar{x}$, de manera que siempre $CV > 0$.

El coeficiente de variación es una medida relativa de variabilidad. En ingeniería se utiliza mucho el coeficiente inverso, $|\bar{x}| / s$, que se conoce como *coeficiente señal-ruido*. Para datos que representen distintas mediciones de una misma magnitud, CV indica la magnitud del error promedio de medición, s, como porcentaje de la cantidad medida.

El coeficiente de variación en datos positivos de una población homogénea es típicamente menor que la unidad. Si este coeficiente es mayor que 1,5, conviene investigar posibles fuentes de heterogeneidad en los datos (medidas con distintos instrumentos; en personas de distinto sexo; en distintos momentos temporales, etc.).

Otras medidas de dispersión

La medida de dispersión que asociamos a la mediana, *Med,* es la mediana de las desviaciones absolutas *(MEDA)* definida por:

$$MEDA = \text{mediana } |x_i - Med| \tag{2.8}$$

que tiene la ventaja, como la mediana, de no verse afectada por datos extremos. A las medidas que tienen esta propiedad las llamaremos *medidas robustas o resistentes*. Si conocemos la mediana y la MEDA de datos no agrupados, sabemos que, al menos, el 50% de los datos está en el intervalo ($Med \pm MEDA$).

Se denomina *rango o recorrido* de una variable la diferencia entre su valor máximo y mínimo. Llamaremos *percentil p* al menor valor superior al p% de los datos. Por ejemplo, si el número de datos es impar, la mediana es el percentil 50. Llamaremos *cuartiles* a aquellos valores que dividen la distribución en cuatro partes iguales. El primer cuartil, Q_1, es por definición igual al percentil 25, el segundo es la mediana y el tercero, Q_3, el percentil 75. Los percentiles y los cuartiles se utilizan para construir medidas de dispersión basadas en los datos ordenados, como el *rango intercuartílico,* que es la diferencia entre los percentiles 75 y 25.

Ejemplo 2.3

Para los datos de la tabla 2.2 hemos visto en los ejemplos 2.1 y 2.2 que la media es 1 y la desviación típica 1,18. Por tanto, el intervalo de la media y dos desviaciones típicas es 1 ± 2.(1,18) = (0; 3,36) y cubre el 94% de los datos, bastante más que el porcentaje mínimo del 75%. Con tres desviaciones típicas obtenemos 1 ± 3.(1,18) = (0; 4,54), que cubre el 99% de los datos, de nuevo, bastante más que el porcentaje mínimo. El coeficiente de variación:

$$CV = \frac{1{,}18}{1} = 1{,}18$$

es un valor ligeramente alto, pero no preocupante. Supongamos que cometemos un error de transcripción y en lugar de seis llamadas como máximo apuntamos por error el valor 16 en la tabla 2.2. Entonces, el lector puede comprobar que la media de los datos sería 1,1 y la desviación típica 1,84, dando lugar a un coeficiente de variación de 1,67, que sería muy indicativo de la presencia de posibles errores en los datos.

La mediana de estos datos es 1, y la MEDA es la mediana de las desviaciones absolutas. La distribución de desviaciones absolutas se obtiene restando 1 a los valores de las variables, tomando el valor absoluto y sumando las frecuencias relativas que dan lugar a la misma desviación absoluta.

2.4 Medidas de asimetría y curtosis

Estas medidas informan sobre dos aspectos importantes de la forma de la distribución: su grado de asimetría y su grado de homogeneidad. Al ser medidas de forma, no dependen de las unidades de medida de los datos.

2.4.1 Coeficiente de asimetría

En un conjunto de datos simétricos respecto a su media $\bar{x}$, la suma $\Sigma\,(x_i - \bar{x})^3$ será nula, mientras que con datos asimétricos esta suma crecerá con la asimetría. Para obtener una medida adimensional, se define el *coeficiente de asimetría* mediante:

$$CA = \frac{\Sigma(x_i - \bar{x})^3}{ns^3} \qquad (2.9)$$

$\lvert x_i - Med \rvert$	0	1	2	3	4	5
fr	0,29	0,44 + 0,26	0,07	0,03	0,00	0,01

La mediana de estas desviaciones es 1, que es la MEDA de esta distribución. Los cuartiles de la distribución de la tabla 2.2 se obtienen suponiendo los datos ordenados y viendo los que corresponden al 25 y 75%. Al 25% corresponde el cero, ya que el 44% de los datos es cero, y el que corresponde al 75% es 2, ya que menor que 2 hay el 73% y mayor que 2 el 11%. El rango intercuartílico será igual a 2.

Para los datos de la tabla 2.3, como la media es 27,65 minutos y la desviación típica 5,08 minutos, el intervalo entre la media y dos desviaciones típicas es 27,65 ± 2.(5,08) = (17,49; 37,81) y cubre hasta algo más de la mitad del intervalo (35,39), lo que supone el 93,5% de los datos, un porcentaje similar al caso anterior. Con tres desviaciones típicas obtenemos 27,65 ± 3.(5,08) = (12,41; 42,89), que cubre el 98,5% de los datos aproximadamente. El coeficiente de variación es

$$CV = \frac{5,08}{27,65} = 0,184$$

que es un valor pequeño. No detallaremos el cálculo de MEDA ni de los cuartiles con datos agrupados porque es tedioso (hay que repartir proporcionalmente las frecuencias con reglas de tres) y conceptualmente no añade nada nuevo. Además, siempre que sea posible, conviene calcular la mediana y los cuartiles antes de agrupar los datos, y los cálculos se realizan normalmente con ordenador.

donde *s* es la desviación típica. El signo del coeficiente de asimetría indica la forma de la distribución. Si este coeficiente es negativo, la distribución se alarga para valores inferiores a la media, como indica la figura 2.10 (a). Si el coeficiente es positivo, la cola de la distribución se extiende para valores superiores a la media, como indica la figura 2.10 (b).

Otra medida de asimetría poco utilizada es:

$$\frac{\bar{x} - mediana}{s}$$

que es también adimensional.

2.4.2 Coeficiente de curtosis

La figura 2.11 presenta cuatro ejemplos de distribuciones de frecuencias simétricas con la misma media (cero) y desviación típica (uno) pero distinta

Figura 2.10 Dos distribuciones asimétricas y sus coeficientes de asimetría

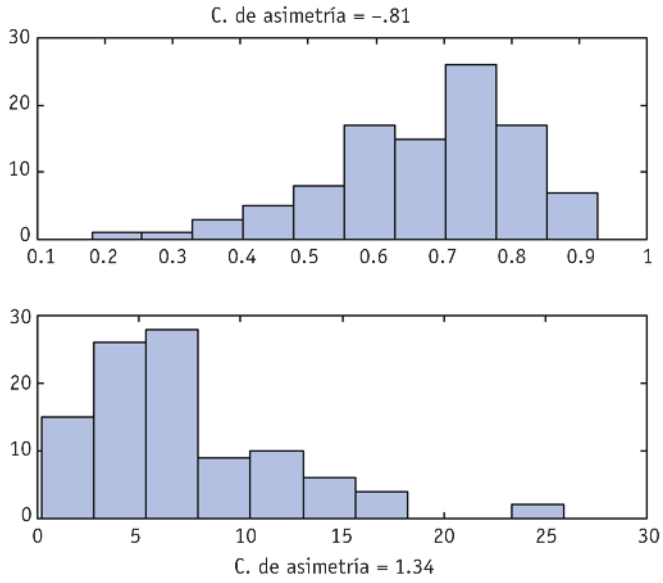
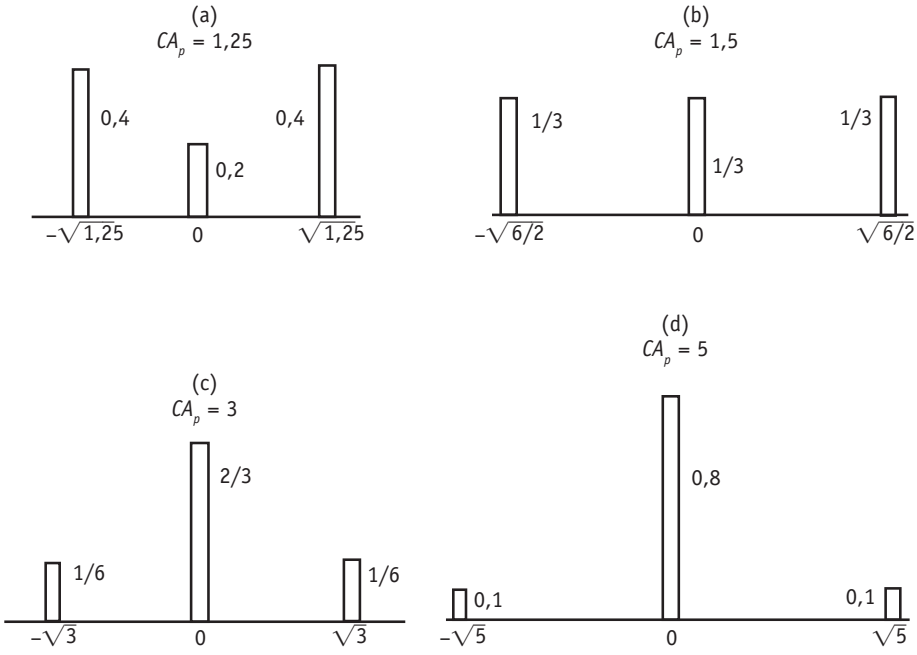


Figura 2.11 Cuatro distribuciones y sus coeficientes de curtosis

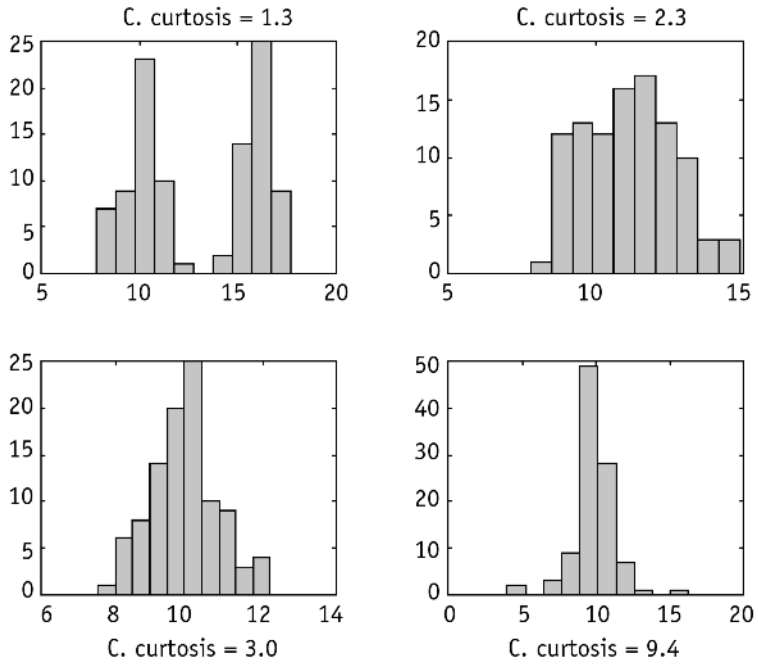


forma: en la primera la frecuencia relativa del valor central es muy baja y normalmente van a observarse valores alejados de la media; en la segunda, la frecuencia relativa de todos los valores es la misma, y también es esperable que aparezcan valores alejados de la media. La situación se invierte en el tercer caso, donde lo frecuente es el valor central, si bien con menor frecuencia pueden aparecer valores alejados. Finalmente, en el cuarto caso pueden aparecer con muy poca frecuencia valores muy extremos. Esta característica, como la frecuencia relativa se reparte entre el centro y los extremos, se denomina *apuntamiento* o *curtosis.* Diremos que las dos primeras distribuciones tienen poco apuntamiento, es decir, poca concentración de probabilidad cerca de la media de la distribución, la tercera un apuntamiento medio y la cuarta un alto apuntamiento, con unos pocos valores extremos. Esta propiedad se mide mediante el *coeficiente de curtosis,* que se define por:

$$CA_p = \frac{\Sigma(x_i - \overline{x})^4}{ns^4} \tag{2.10}$$

donde s es la desviación típica de los datos. Este coeficiente es siempre mayor o igual que uno.

Figura 2.12 Cuatro distribuciones y su coeficiente de curtosis



El coeficiente de curtosis es importante porque nos informa respecto a la heterogeneidad de la distribución. La figura 2.12 presenta cuatro distribuciones que ilustran con datos reales la situación esquematizada en la figura 2.11. Las cuatro corresponden a los tiempos de servicio requeridos por distintos clientes en distintos servicios. En el primer caso el apuntamiento de la distribución es 1,25, y este bajo valor es indicativo de una distribución muy heterogénea. La distribución que observamos es una mezcla de los tiempos de servicio de dos tipos de clientes que se observa están aproximadamente repartidos al 50%. En el segundo caso tenemos varios tipos de clientes, pero sus tiempos de servicio son más próximos, dando lugar a una distribución menos heterogénea con curtosis 1,69. La tercera distribución representa el tiempo de servicio cuando los clientes son homogéneos y la curtosis es igual a 3. En la cuarta los clientes son homogéneos, pero existen de vez en cuando valores extremos que requieren un valor o muy alto o muy bajo. Estos clientes son atípicos, y dan lugar a un coeficiente de curtosis muy alto, de 9,4. El coeficiente de curtosis nos informa de la posible heterogeneidad en los datos. Si es muy bajo (menor de 2), indica una distribución mezclada; si es muy alto (mayor de 6), indica la presencia de valores extremos atípicos.

En la fórmula anterior, para calcular el coeficiente de apuntamiento se ha supuesto que los datos están sin agrupar. Para datos agrupados el numerador va extendido como siempre a los valores distintos y cada término se multiplica por su frecuente relativa. Por ejemplo, el CA_p de la distribución de la figura 2.11 (a) será:

$$\sum (x_j - \bar{x})^4 \, fr(x_j) = (-\sqrt{1{,}25})^4 \cdot 0{,}4 + (-\sqrt{1{,}25})^4 \cdot 0{,}4 = 1{,}25$$

Como para esta distribución $s = 1$, el *CAp* será igual a 1,25. Si suponemos una distribución con valores posibles (–1 y 1) con frecuencias relativas (0,5, 0,5), la desviación típica es 1 y

$$\sum (x_j - \bar{x})^4 \, fr(x_j) = (-1)^4 \cdot 0{,}5 + (1)^4 \cdot 0{,}5 = 1$$

y el coeficiente de apuntamiento será también uno, que es el valor mínimo de coeficiente.

2.4.3 Otras medidas características

Para describir otros aspectos relevantes de la distribución de frecuencias se utilizan los momentos de la distribución. Definimos *momento de orden k* respecto al origen como:

Ejemplo 2.4

Calcularemos los coeficientes de asimetría y apuntamiento para los datos de las tablas 2.2 y 2.3. Comenzando con los datos de la tabla 2.2:

$$\sum (x_j - 1)^3 \, fr(x_j) =$$

$$= (-1)^3 \, 0{,}44 + (1)^3 \, 0{,}16 + (2)^3 \, 0{,}07 + (3)^3 \, 0{,}03 + (5)^3 \, 0{,}01 = 2{,}34$$

$$\sum (x_j - 1)^4 \, fr(x_j) =$$

$$= (-1)^4 \, 0{,}44 + (1)^4 \, 0{,}16 + (2)^4 \, 0{,}07 + (3)^3 \, 0{,}03 + (5)^4 \, 0{,}01 = 10{,}40$$

$$CA = \frac{2{,}34}{1{,}18^3} = 1{,}42; \qquad CA_p = \frac{10{,}4}{(1{,}4)^2} = 5{,}31$$

El coeficiente de asimetría positivo nos indica que la distribución se alarga para valores mayores que la media. El coeficiente de apuntamiento es alto, pero no tanto para hacernos concluir que deben existir valores extremos en la distribución.

Para los datos de la tabla 2.3, como la media es 27,65 (ejemplo 2.1), y la desviación típica 5,08, tenemos que:

$$\sum (x_j - 27{,}65)^3 \, fr(x_j) =$$

$$= (22 - 27{,}65)^3 \, 0{,}3 + (27 - 27{,}65)^3 \, 0{,}4 + (32 - 27{,}65)^3 \, 0{,}2 + \\ + (37 - 27{,}65)^3 \, 0{,}07 + (42 - 27{,}65)^3 \, 0{,}03 = 108{,}11$$

$$\sum (x_j - 27{,}65)^4 \, fr(x_j) =$$

$$= (22 - 27{,}65)^4 \, 0{,}3 + (27 - 27{,}65)^4 \, 0{,}4 + (32 - 27{,}65)^4 \, 0{,}2 + \\ + (37 - 27{,}65)^4 \, 0{,}07 + (42 - 27{,}65)^4 \, 0{,}03 = 2184{,}5$$

$$CA = \frac{108{,}11}{(5{,}08)^3} = 0{,}82; \qquad CA_p = \frac{2184{,}5}{(25{,}82)^2} = 3{,}27$$

El signo de la asimetría nos indica que la distribución no es simétrica y se alarga hacia valores mayores que la media. El coeficiente de curtosis toma un valor medio.

$$m_k = \frac{\Sigma\, x_j^4}{n}$$

Por tanto, $\bar{x}$ es el momento de orden 1 respecto al origen. Los momentos respecto a la media se definen por:

$$\mu_k = \frac{\Sigma\, (x_j - \bar{x})^k}{n}$$

La varianza es el segundo momento respecto a la media. La medida adimensional de apuntamiento (2.10) suele escribirse:

$$CA_p = \frac{\mu_4}{s^4}$$

2.5 Datos atípicos y diagramas de caja

2.5.1 Datos atípicos

Es muy frecuente que los datos presenten observaciones que contienen errores de medida o de transcripción o que son heterogéneas con el resto porque se han obtenido en circunstancias distintas. Llamaremos datos atípicos a estas observaciones generadas de forma distinta al resto de los datos. Los análisis efectuados sobre datos recogidos en condiciones de estrecho control revelan que es frecuente que aparezcan entre un 1 y un 3% de observaciones atípicas en la muestra. Cuando los datos se han recogido sin un cuidado especial, la proporción de datos atípicos puede llegar al 5% y ser incluso mayor.

Los datos atípicos se identifican fácilmente con un histograma o diagrama de barras de los datos, porque aparecerán separados del resto de la distribución. Sin embargo, en el análisis automático de muchas variables es conveniente tener reglas simples para detectarlos. Un criterio simple es considerar sospechosas aquellas observaciones alejadas de la media más de tres desviaciones típicas. La justificación de esta regla es que, como hemos visto, entre la media y tres desviaciones típicas debe estar al menos el 89% de los datos. Un problema con esta regla es que si existen varios valores atípicos muy grandes que distorsionan la media y la desviación típica, es posible que los datos atípicos no sean identificados, como veremos en el ejemplo 2.5. Una regla mejor es utilizar valores de centralización y dispersión que estén poco afectados por valores atípicos, como la mediana y la Meda. La regla para identificar atípicos es

$$x > Med \pm 4{,}5 \times Meda$$

es decir, consideramos sospechosas observaciones que se alejan de la mediana más de cuatro veces y media la Meda.

Este criterio es simple y se utiliza mucho, pero presenta el inconveniente de no tener en cuenta la asimetría de la distribución. Un criterio más elaborado es partir de los tres cuartiles que dividen los datos en cuatro partes iguales y considerar extremos aquellos valores que se alejan una cantidad definida por la izquierda del primer cuartil, Q_1, o por la derecha del tercer cuartil, Q_3. Como medida de dispersión en lugar de la Meda se utiliza entonces el rango intercuartílico ($Q_3 - Q_1$), y se consideran atípicas aquellas observaciones que son menores de

$$x < Q_1 - 1{,}5(Q_3 - Q_1)$$

o son mayores de

$$x > Q_3 + 1{,}5(Q_3 - Q_1)$$

Los datos identificados como atípicos o sospechosos deben comprobarse para ver si es posible encontrar la causa de la heterogeneidad. Cuando no se encuentre un error, hay que sospechar que sobre esa observación ha actuado alguna causa que no ha estado actuando en el resto de las observaciones. Por ejemplo, alguna variable que afecta a la que observamos ha tomado un valor distinto y es responsable del cambio observado. El descubrimiento de esta variable insospechada puede ser el resultado más importante del estudio descriptivo. Muchos descubrimientos científicos importantes y muchas patentes industriales han surgido de la investigación para determinar las razones de un dato anómalo. En último caso, cuando la observación sea muy extrema, entendiendo por ello más alejada de la mediana de 8 veces la Meda o situada fuera del intervalo [$Q_1 - 3(Q_3 - Q_1)$; $Q_3 + 3(Q_3 - Q_1)$] conviene, aunque no se encuentre la causa, descartarla del análisis.

2.5.2 Diagrama de caja

El diagrama de caja es una representación semigráfica de una distribución construida para mostrar sus características principales y señalar los posibles datos atípicos. Se diferencia de las representaciones gráficas anteriores en que está especialmente pensada para identificar los valores atípicos que pueden afectar a todo el análisis posterior. Se construye como sigue:

1. Ordenar los datos de la muestra y obtener el valor mínimo, el máximo y los tres cuartiles Q_1, Q_2, Q_3.

2. Dibujar un rectángulo cuyos extremos son Q_1 y Q_3 e indicar la posición de la mediana (Q_2) mediante una línea.
3. Calcular unos límites admisibles superior e inferior que van a servir para identificar los valores atípicos. Estos límites se calculan con:

$$\begin{aligned} LI &= Q_1 - 1{,}5(Q_3 - Q_1) \\ LS &= Q_3 + 1{,}5(Q_3 - Q_1) \end{aligned}$$

4. Considerar como valores atípicos los situados fuera del intervalo (*LI, LS*).
5. Dibujar una línea que vaya desde cada extremo del rectángulo central hasta el valor más alejado no atípico, es decir, que está dentro del intervalo (*LI, LS*).
6. Identificar todos los datos que están fuera del intervalo (*LI, LS*), marcándolos como atípicos.

Los diagramas de caja son especialmente útiles para comparar la distribución de una variable en distintas poblaciones.

Ejemplo 2.5

La tasa de incremento de los precios al consumo en 1985 de los 24 países de la OCDE fue (con un asterisco aparecen los miembros en ese momento de la Unión Europea):

Alem.	Austr.	Austria	Bélgica	Canadá	Dinam.	España	EE.UU.	Finland.
2,2(*)	7,6	2,9	4,6(*)	4,1	3,9(*)	7,4(*)	3,2	5,1

Franc.	Grecia	Holand.	Irland.	Island.	Italia	Japón	Luxem.	Noruega
5,3(*)	20,1(*)	2,3(*)	5,5(*)	32,7	9,1(*)	1,7	3,2(*)	5,8

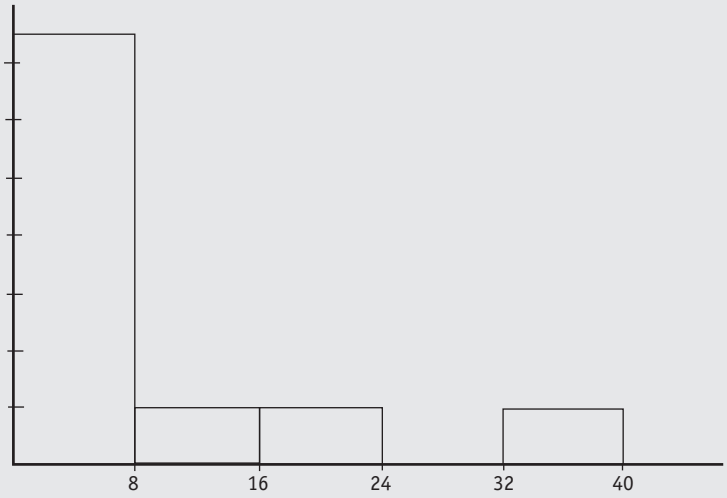
N. Zel.	Portug.	G. Bret.	Suecia	Suiza	Turquía
16,3	15,9(*)	5,9(*)	6,7	3,4	40,5

Agrupando los datos en cinco clases (aplicando la regla de $\sqrt{n}$), se obtiene la tabla:

Intervalo	0 a 8	8 a 16	16 a 24	24 a 32	32 a 40
Frecuencia	18	2	2	0	2
Frecuencia relat.	0,75	0,08	0,08	0	0,08

El histograma de estos datos se indica en la figura 2.13, y muestra que los datos son muy asimétricos: la mayoría de los países tienen una inflación entre 1,7 y 8,35, y unos pocos una inflación muy alta.

Figura 2.13 Histograma para los datos de la inflación de la OCDE



Las medidas características para estos datos, utilizando los datos originales, son:

$$\bar{x} = \frac{2{,}2 + 7{,}6 + \ldots + 40{,}5}{24} = 8{,}98$$

mediana = 5,4
desviación típica = $[(2{,}2 - 8{,}98)^2/24 + \ldots + (40{,}5 - 8{,}98)^2/24)]^{1/2} = 9{,}78$
$CA = (9{,}78)^{-3}\,[(2{,}2 - 8{,}98)^3/24 + \ldots + (40{,}5 - 8{,}98)^3/24] = 1{,}98$
$CA_p = (9{,}78)^{-4}\,[(2{,}2 - 8{,}98)^4/24 + \ldots + (40{,}5 - 8{,}98)4^3/24] = 6{,}10$

La meda es la mediana de las desviaciones resultantes de restar a cada dato la mediana, con lo que se obtiene el conjunto (3,2; 2,2; 2,5; ...; 1,3; 2,0; 35,1), cuya mediana es 2,2.

Si aplicamos los criterios para encontrar atípicos, con la media y la desviación típica, se considerarán sospechosos los situados fuera del intervalo

$$8{,}98 \pm 3 \times 9{,}78 = (0{,}38{,}32)$$

y sólo Turquía está fuera de ese rango. Sin embargo, el histograma muestra claramente que los países con inflación mayor que 32 son claramente distintos del resto.

Si utilizamos estadísticos robustos, como la mediana y la meda, tenemos que el intervalo será:

$$5{,}4 \pm 4{,}5 \times 2{,}22 = (0{,}15{,}39)$$

y cinco países salen fuera de este intervalo y se identifican como sospechosos. Vamos a comparar estos resultados con los obtenidos con el diagrama de caja. Para calcularlo, ordenamos los datos de menor a mayor:

(1,7; 2,2; 2,3; 2,9; 3,2; 3,2; 3,4; 3,9; 4,1; 4,6; 5,1; 5,3; 5,5; 5,8; 5,9; 6,7; 7,4; 7,6; 9,1; 15,9; 16,3; 20,1; 32,7; 40,5)

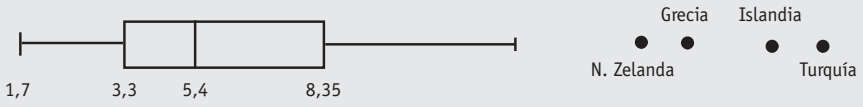
Como en la posición 12 y 13 están los valores 5,3 y 5,5, la mediana será su media, 5.4. Análogamente, los cuartiles serán:

$$Q_1 = \frac{3{,}2 + 3{,}4}{2} = 3{,}3$$
$$Q_3 = \frac{7{,}6 + 9{,}1}{2} = 8{,}35$$

Por tanto, los límites admisibles son:

$$LI = 3{,}3 - 1{,}5\ (8{,}35 - 3{,}3) = -4{,}275$$
$$LS = 8{,}35 + 1{,}5\ (8{,}35 - 3{,}3) = 15{,}92$$

Como todos los valores son superiores al límite inferior, la línea inferior del diagrama de caja deberá llegar hasta el valor mínimo y no hay atípicos en esa dirección. Por el contrario, el valor más alto incluido en el intervalo (0; 15,92) es Portugal, con 15,9, que será el límite de la línea superior del diagrama de caja. Los otros cuatro países deben considerarse atípicos y representarse en el gráfico para su identificación.



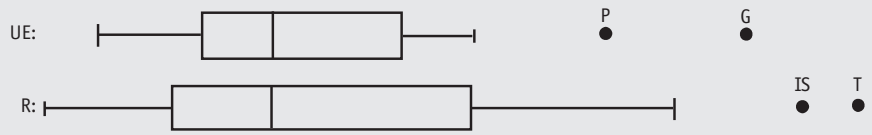
En este caso vemos que la presencia de valores muy heterogéneos hace que la información del diagrama de caja sea más útil que la del histograma.

Supongamos ahora que comparamos la distribución de los países que eran entonces miembros de la Unión Europea (marcados con un asterisco) con el resto. Tenemos:

UE: (2,2; 2,3; 3,2; 3,9; 4,6; 5,3; 5,5; 5,9; 7,4; 9,1; 15,9; 20,1)
Resto: (1,7; 2,9; 3,2; 3,4; 4,1; 5,1; 5,8; 6,7; 7,6; 16,3; 32,7; 40,5)

Para los países de la UE la mediana es 5,4[(5,3 + 5,5)/2] y los cuartiles (3,2 + 3,9)/2 = 3,55; (7,4 + 9,1)/2 = 8,25. Para el resto, estos valores son Q_1 = 3,3; Q_2 = 5,5; Q_3 = 11,95. Los intervalos de admisibilidad son (3,55 – 3,525; 8,25 + 3,525) para la UE (ya que 1,5 (8,25 – 3,55)/2 = 3,525) y (3,3 – 6,427; 11,95 + 6,487) para el resto. La figura 2.14 muestra los diagramas de caja para ambos colectivos. Se observa que en ambos hay dos valores atípicos y que la variabilidad es mucho más baja entre los países de la UE.

Figura 2.14 Diagramas de caja para la inflación en la OCDE



2.6 Transformaciones

2.6.1 Transformaciones lineales

El objetivo central de la descripción de los datos es obtener una visión tan clara y simple como sea posible, y las unidades de medida de la variable deben escogerse con este criterio. Por ejemplo, si x es la estatura en metros y se han observado los valores 1,75; 1,68; 1,80; ...; con 1,65 como menor valor, la transformación

$$y = 100\ (x - 1{,}65)$$

conduce al conjunto de datos: 10; 3; 15; ...; de tratamiento más simple. En general en la descripción inicial de los datos *conviene representarlos con únicamente dos o tres dígitos,* escogiendo apropiadamente las unidades. Esto equivale a efectuar una transformación lineal:

$$y = a + bx$$

Las medidas características de la variable original, x, se obtienen fácilmente a partir de las calculadas para la transformada, y, ya que es inmediato comprobar que:

$$\bar{y} = \frac{\Sigma y}{n} = \frac{\Sigma(a + bx)}{n} = a + b\bar{x}$$
$$s_y = |b| s_x$$

y los coeficientes de asimetría y curtosis no se alteran, al ser adimensionales. Esta transformación es importante con datos con muchos dígitos comunes, ya que entonces, además de una representación más clara, aumentamos la precisión de los cálculos realizados con una máquina de calcular o un ordenador personal.

2.6.2 Transformaciones no lineales

En muchos problemas el fenómeno estudiado puede medirse mediante variables relacionadas no linealmente entre sí. Por ejemplo, el consumo de gasolina de un automóvil se expresa en Europa en litros cada 100 km (x) y en Estados Unidos en km recorridos con 1 litro (o galón) de gasolina (y). La relación entre ambas medidas es no lineal, ya que $y = 100/x$. Como segundo ejemplo, se desea comparar el crecimiento del consumo de energía en distintos países. Una posibilidad es estudiar las diferencias $C_t - C_{t-1}$, pero en general resulta más relevante considerar las diferencias relativas $(C_t - C_{t-1})/C_{t-1}$ o $(C_t - C_{t-1})/C_t$. Si expresamos la variable en logaritmos, su crecimiento en dicha escala es una buena medida del crecimiento relativo, ya que:

$$\ln C_t - \ln C_{t-1} = \ln \frac{C_t}{C_{t-1}} = \ln \left(1 + \frac{C_t - C_{t-1}}{C_{t-1}} \right) \simeq \frac{C_t - C_{t-1}}{C_{t-1}}$$

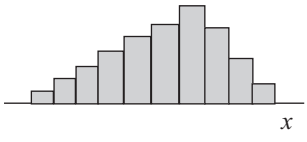
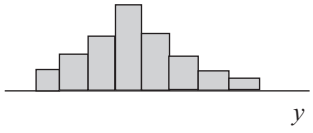
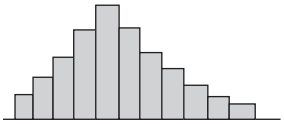
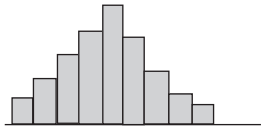
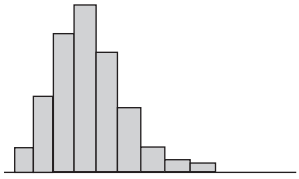
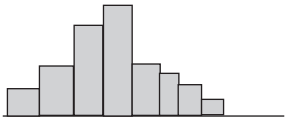
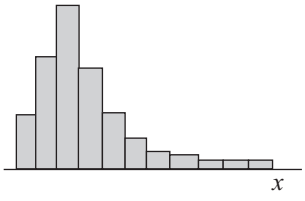
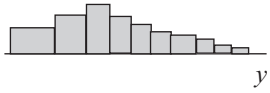
utilizando que $\ln (1 + x)$ es aproximadamente x, si x es pequeño. Además, es fácil demostrar que, supuesto $C_t \geqslant C_{t-1}$:

$$\frac{C_t - C_{t-1}}{C_t} \leqslant \ln \frac{C_t}{C_{t-1}} \leqslant \frac{C_t - C_{t-1}}{C_{t-1}}$$

y las diferencias de las variables en logaritmos son una medida promedio de las dos formas posibles de medir el crecimiento relativo.

Como regla general, conviene escoger aquella transformación que conduzca a una representación lo más simple posible. Las distribuciones simétricas respecto a la media son más simples que las asimétricas, ya que: (1) la media, la mediana y la moda coinciden; (2) el coeficiente de asimetría y to-

Cuadro 2.1 El efecto de las transformaciones

Histograma inicial	Transformación	Histograma transformado
	$y = x^2$	
	$y = \sqrt{x}$	
	$y = \ln x$	
	$y = \frac{1}{x}$	

dos los momentos respecto a la media de orden impar son nulos. Por tanto, cuando exista una transformación, $h(x)$, tal que la nueva variable $y = h(x)$ tenga distribución simétrica, es conveniente trabajar con esta variable transformada.

Las transformaciones más utilizadas se resumen en el cuadro 2.1. La transformación $y = x^2$ comprime la escala para valores pequeños y la expande para valores altos. Es útil para conseguir simetría en distribuciones con coeficiente de asimetría negativo. Por el contrario, las tres transformaciones $\sqrt{x}$, $\ln x$ y $1/x$ comprimen los valores altos y expenden los bajos, produciendo además este efecto en orden creciente (menos $\sqrt{x}$, más $\ln x$ y más todavía $1/x$).

La transformación más utilizada es el logaritmo. Muchas distribuciones que describen el tamaño de las cosas (ciudades en el mundo, tamaño de

Cuadro 2.2 Ejemplo de aplicación del logaritmo

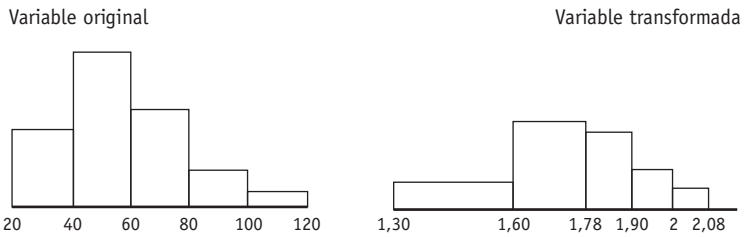
1. Distribución original — **Distribución transformada**

x	*fr*	Como	y	*fr*
20-40	0,20	log 20 = 1,30	1,30-1,60	0,20
40-60	0,40	log 40 = 1,60	1,60-1,78	0,40
60-80	0,25	log 60 = 1,78	1,78-1,90	0,25
80-100	0,10	log 80 = 1,90	1,90-2,00	0,10
100-120	0,05	log 100 = 2,00	2,00-2,08	0,05
		log 120 = 2,08		

2. Cálculo de las alturas en el nuevo histograma

Clase (1)	Longitud (2)	Frecuencia (3)	Altura (3)/(2)
1,30-1,60	0,30	0,20	0,20/0,30 = 0,67
1,60-1,78	0,18	0,40	0,40/0,18 = 2,22
1,78-1,90	0,12	0,25	0,25/0,12 = 2,08
1,90-2,00	0,10	0,10	0,10/0,10 = 1
2,00-2,08	0,08	0,05	0,05/0,08 = 0,63

3. Histogramas



empresas, distribución de rentas, consumo de electricidad, etc.) son aproximadamente simétricas al expresar la variable en logaritmos.

Siempre que sea posible es conveniente transformar los datos originales (x) y construir la nueva distribución de frecuencias a partir de los valores transformados $y = h(x)$. A veces esto no es posible y tenemos que trabajar con la distribución agrupada de x. El cuadro 2.2 presenta un ejemplo de la aplicación de la transformación en esos casos.

El efecto de una transformación depende del rango de los datos, ya que cualquier transformación es aproximadamente lineal en un rango suficientemente pequeño. Como regla general, si el cociente entre el valor máximo y el mínimo es pequeño (menor que dos), la transformación no variará apreciablemente la forma de la distribución, mientras que cuando este cociente sea grande (mayor de 10), el efecto será muy acusado.

Relación entre las medidas características de los datos y sus transformadas

Las medidas basadas en el orden de los datos se mantienen para cualquier transformación monótona, $y = h(x)$, que conserve el orden de los datos, como x^2 (para datos positivos), $\sqrt{x}$ o $\ln x$. Es decir, si

$$x_1 > x_2 \rightarrow h(x_1) > h(x_2) \rightarrow y_1 > y_2$$

por ejemplo, para el logaritmo:

$$\begin{aligned} \textit{Mediana}\ (y) &= \log\ [\textit{mediana}(x)] \\ \textit{Percentil}\ p(y) &= \log\ [\textit{percentil}\ p(x)] \end{aligned}$$

Sin embargo, la media y desviación típica de los datos con una tranformación no lineal no pueden deducirse fácilmente a partir de las originales. Es decir, si $y = \ln x$, en general:

$$\begin{aligned} \bar{y} &\neq \ln \bar{x} \\ s_y &\neq \ln s_x \end{aligned}$$

En el caso de la transformación logarítmica, la media de los datos transformados

$$\bar{y} = \frac{1}{n} (\ln x_1 + ... + \ln x_n)$$

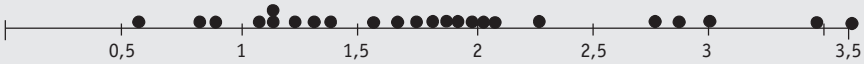
equivale a un valor g en las unidades originales de los datos, donde $\bar{y} = \ln g$, que verifica:

$$g = (x_1 \cdot x_2 \cdot ... \, x_n)^{1/n}$$

que se conoce como media geométrica.

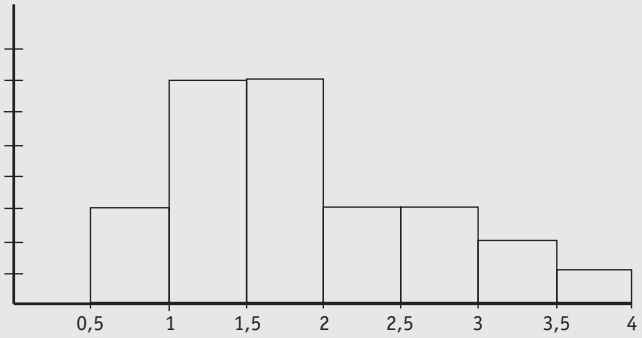
Ejemplo 2.6

Vamos a transformar los datos de inflación del ejemplo 2.5, ya que en las unidades originales presentan gran heterogeneidad. Tomando logaritmos neperianos en los datos, se obtienen los siguientes valores transformados:



De nuevo la mayoría de los datos está en el intervalo (0,5-2,3), que corresponde a una inflación menor del 10%, con un grupo de países heterogéneos con el resto.

Figura 2.15 Histograma para los datos en logaritmos de la inflación de la OCDE



Ejercicios 2

2.1. Calcule la media, mediana y desviación típica de los datos siguientes: 28, 22, 35, 42, 44, 53, 58, 41, 40, 32, 31, 38, 37, 61, 25, 35.

a) Directamente.

b) Agrupando en 5 clases de longitud 10 cm (20-30; 30-40; etc.) y utilizando las fórmulas para datos agrupados.

c) Construya un diagrama de tallo y hojas y un histograma de estos datos.

2.2. Construya una distribución de frecuencias del número de vehículos que pasan por un punto de circulación en un intervalo de un minuto. Calcule distintas medidas de centralización y dispersión y comente su significado.

El histograma para los datos agrupados, tomando intervalos de 0,5 para simplificar, se presenta en la figura 2.17.

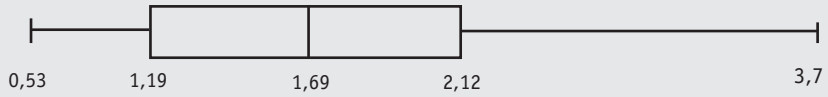
Las medidas características calculadas directamente con los datos transformados son:

$$\bar{y} = 1{,}81; \quad s_y = 0{,}83; \quad CA = 0{,}69; \quad CA_p = 2{,}64$$

Vamos a construir el diagrama de caja. Como el logaritmo conserva el orden, la mediana y los cuartiles de los datos transformados serán los transformados de los cuartiles originales (con número impar de datos exactamente, y aproximadamente cuando este número sea par). Entonces

$$Q_1 = 1{,}19;\ Q_2 = 1{,}69;\ Q_3 = 2{,}12$$
$$LI = 1{,}19 - 1{,}5(2{,}12 - 1{,}19) = -0{,}67$$
$$LS = 2{,}12 + 1{,}5(2{,}12 - 1{,}19) = 3{,}98$$

y ahora ningún país es atípico. El diagrama de caja será:



y es considerablemente más simétrico.

La conclusión de este ejercicio es que un conjunto de datos puede parecer muy heterogéneo en una escala y homogéneo en otra, transformando adecuadamente los datos.

Método: Elija un punto de tráfico denso y uniforme y con un cronómetro cuente cuántos vehículos (defina de qué tipo va a considerar) pasan en intervalos de un minuto. Haga la distribución de frecuencias de los datos y un diagrama de barras.

2.3. Construya una distribución de frecuencias de las siguientes variables de las provincias españolas: tamaño, población, densidad de población, natalidad, matrimonios y cualquier otra de su interés. Los datos puede encontrarlos en el banco de datos del Instituto Nacional de Estadística, http://www.ine.es.
Repita el análisis por comunidades autónomas. ¿Qué conclusiones se obtienen? ¿Mejora la simetría con alguna transformación?

2.4. Estudie la distribución del tamaño de la palabra en distintos idiomas, calculando medidas descriptivas de esta distribución.

Método: Elija novelas en tres idiomas (por ejemplo, castellano, francés e inglés) y cuente en distintas páginas la frecuencia de aparición de palabras de una, dos tres... letras. Haga un diagrama de barras, calcule la media y la desviación típica e interprete los resultados.

2.5. Estudie la distribución del tiempo que tarda en desplazarse a clase cada mañana. Realice histogramas para distintos días de la semana y horas del día. ¿Qué conclusiones pueden obtenerse?

2.6. Encuentre el valor a que minimiza $\Sigma(x_i - a)^2$. Conclusiones.

2.7. Demuestre que si multiplicamos todos los valores de una variable por k, la media y la desviación típica quedarán multiplicadas también por k ($k > 0$).

2.8. Estudie la distribución de la longitud de las canciones de su autor favorito. Construya un histograma y calcule las medidas descriptivas estudiadas mediante un programa de ordenador. Compare la longitud de las canciones en períodos distintos mediante gráficos *box-plot* para ver si hay evidencia de que la longitud ha variado con el tiempo.
Como ejemplo, las longitudes del CD de J. Sabina, «19 días y 500 noches», publicado en 1999, son, en minutos y segundos: 6:69, 4:45; 6:37; 4:15; 5:35; 7:11; 4:50; 3:32; 4:37; 7:29; 4:52; 8:41; 4:42, y las longitudes del CD de 1993, «Querido Sabina», del mismo autor, fueron: 3:40; 3:39; 4:07; 4:27; 5:46; 6:06; 2:39; 5:43; 4:15; 4:09; 5:15. Si trasladamos todas estas medidas a minutos, mediante la fórmula: tiempo en minutos = minutos + $\frac{segundos}{60}$, las longitudes del primer CD son 6.82, 4.75, 6.62, 4.25, 5.58, 7.18, 4.83, 3.53, 4.62, 7.48, 4.87, 8.68, 4.70, y las del segundo, 3.67, 3.65, 4.12, 4.45, 5.77, 6.10, 2.65, 5.72, 4.25, 4.15, 5.25.

2.9. Demuestre que la media aritmética de la variable z obtenida sumando los datos de otras dos variables, x e y, es la suma de las medias aritméticas de éstas.

2.10. Demuestre que si construimos una variable z mezclando n_1 valores de x y n_2 de y, la media de z es:

$$\overline{z} = \frac{n_1}{n_1 + n_2}\overline{x} + \frac{n_2}{n_1 + n_2}\overline{y}$$

siendo $\overline{x}$, $\overline{y}$ las medias de las variables iniciales.

2.11. Si $z = x + y$, demostrar que la varianza de z puede ser mayor, menor o igual que la suma de las varianzas de los sumandos.

2.12. Demostrar que los momentos respecto a la media de orden 3 están relacionados con los momentos respecto al origen por la expresión:

$$\mu_3 = m_3 - 3m_2m_1 + 2m_1^3$$

2.13. La variable x toma los valores 1, 2, 3, 4 y 5 con frecuencia relativa 0,2 para todos ellos. La y es también simétrica con valores ($3 - a$; 2, 9; 3; 3, 1; $3 + a$) y frecuencias relativas respectivas (0,05; 0,2; 0,5; 0,2; 0,05). Se pide:

a) Encontrar el valor de a, para el que ambas distribuciones tienen la misma varianza.
b) Calcular el coeficiente de apuntamiento para x e y, suponiendo el valor anterior de a.

2.14. En una distribución de frecuencias podemos asociar la frecuencia de cada dato x_i a la masa situada en dicho punto. Entonces la media $\bar{x} = \Sigma x_i f_i / \Sigma f_i$ corresponde al centro de gravedad de las observaciones. ¿Qué podríamos asociar entonces al momento de inercia?

2.15. En 1879 Michelson obtuvo los siguientes valores para la velocidad de la luz en el aire (damos los resultados restando 299.000 a los datos originales, en km/seg., para facilitar su manejo): 850, 740, 900, 1.070, 930, 850, 950, 980, 980, 880, 1.000, 980, 930, 650, 760. En 1882 Newcomb, utilizando otro procedimiento, obtuvo (restando de nuevo 299.000): 883, 816, 778, 796, 682, 711, 611, 599, 1.051, 781, 578, 796, 774, 820, 772. Se pide:

a) Construya diagramas de tallo y hojas para ambas distribuciones.
b) Calcule medias y desviaciones típicas.
c) ¿Qué conclusiones pueden extraerse?

2.16. Se define la media geométrica por:

$$G = (x_1 \ldots x_n)^{1/n}$$

y la media armónica por:

$$H = \left(\frac{1}{n} \sum \frac{1}{x_i} \right)^{-1}$$

Se pide:

a) Explicar la relación entre la media geométrica y la media de la variable en logaritmos.
b) Lo mismo entre H y la transformación $y = x^{-1}$.

2.17. Demostrar que si una cantidad crece durante k períodos con tasas de crecimiento r_1, ... r_k, donde $r_i = V_i/V_{i-1}$, siendo V_i el valor al final del período i, la tasa media de crecimiento durante el período es la media geométrica de las tasas parciales.

2.18. Demostrar que la varianza puede calcularse mediante $\Sigma \Sigma (x_i - x_j)^2/2n^2$.

2.7 Resumen del capítulo y consejos de cálculo

En este capítulo hemos estudiado, primero, cómo describir una variable estadística, y segundo, cómo describir la interdependencia de un conjunto de variables.

La herramienta principal en la descripción de una variable es su distribución de frecuencias, que es la tabla o gráfico que representa los valores observados y sus frecuencias relativas. Esta distribución refleja dos aspectos fundamentales:

1. La homogeneidad de los datos. Los datos son heterogéneos cuando la distribución muestra varias modas, una gran dispersión o la presencia de valores atípicos muy alejados del resto.
2. La forma de distribución.

Cuando dispongamos de datos homogéneos podemos calcular medidas resumen de su distribución. Las más importantes son la media y la mediana, la desviación típica, el coeficiente de asimetría y el coeficiente de apuntamiento. Estos cinco parámetros resumen concisamente las características de la distribución.

Una segunda idea importante es que conviene seleccionar la escala de medida de los datos para obtener una representación lo más simple posible. Esto incluye tanto transformaciones lineales (por ejemplo, desviaciones a un valor de referencia) como no lineales (logaritmo, raíz...) que conduzcan a distribuciones homogéneas y simétricas.

Estos mismos principios se aplican al estudio conjunto de varias variables, como veremos en el capítulo siguiente. Entonces, además de las medidas características de las variables individuales (distribuciones marginales), conviene incluir las medidas de la relación lineal entre las variables, como el coeficiente de correlación. La medida global de variabilidad utilizada es la matriz de varianzas y covarianzas.

El cuadro 2.3 resume las fórmulas principales introducidas en este capítulo.

Cuadro 2.3 Fórmulas principales del capítulo 2

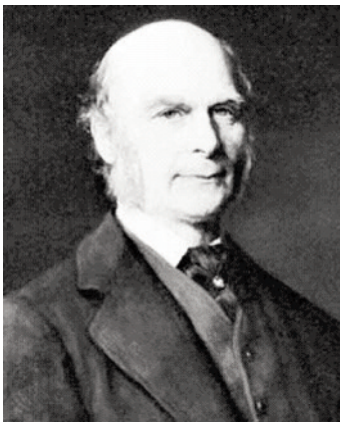
	Datos sin agrupar	Datos agrupados
Frecuencia relativa [$fr\ (x_j)$]	$1/n$	$\dfrac{\text{(n.º de datos con } x_j)}{\text{n.º total de datos}}$
Media ($\bar{x}$)	$\Sigma\, x_i/n$	$\Sigma x_j\, fr(x_j)$
Desviación típica (s)	$\sqrt{\Sigma\,(x_i-\bar{x})^2/n}$	$\sqrt{\Sigma\,(x_i-\bar{x})^2 fr\ (x_j)}$
Desigualdad de Tchebychev	$fr(\lvert x_i-\bar{x}\rvert \le k) \ge 1-1/k^2$	
Coeficiente de variación (CV)	$s/\lvert\bar{x}\rvert$	
Coeficiente de asimetría (CA)	$\Sigma(x_i-\bar{x})^3/ns^3$	$\Sigma\left(\dfrac{x_j-\bar{x}}{s}\right)^3 fr(x_j)$
Coeficiente de apuntamiento (CA_p)	$\Sigma(x_i-\bar{x})^4/ns^4$	$\Sigma\left(\dfrac{x_j-\bar{x}}{s}\right)^4 fr(x_j)$
Diagrama de caja	$L(S/I) = Q_1 \pm 1{,}5\ (Q_3 - Q_1)$	
Transformaciones ($y = a + bx$)	$\bar{y} = a + b\bar{x}$; $\quad s_y = \lvert b\rvert s_x$	

Las medidas estudiadas en este capítulo pueden obtenerse con cualquier programa informático. Por ejemplo, Excel proporciona como funciones la media, desviación típica y coeficientes de asimetría y curtosis de los datos y permite hacer diagramas de barras y pictogramas. Statgraphics, Minitab y los restantes programas estadísticos permiten calcular además histogramas y *box-plots* (bajo el nombre *Box-and-Whisker plot*). En algunos de estos programas el valor que proporcionan del coeficiente de curtosis es el resultado de aplicar la fórmula del cuadro 2.3 y después restar 3, que se toma como valor de referencia. Recomendamos al lector que se ejercite en la interpretación de estos coeficientes calculándolos con un programa informático para distintos conjuntos de datos que puede encontrar fácilmente en Internet.

2.8 Lecturas recomendadas

Tukey (1977) y Ehrenberg (1986) son excelentes introducciones al análisis descriptivo y exploratorio de datos. Otras buenas introducciones son Vallemen y Hoaglin (1981), Hoaglin y otros (1985, 2000) y Mosteller y otros (1983). Chambers *et al.* (1983), Tufte (2001) y Cleveland (1993, 1994) analizan las representaciones gráficas en estadística.

3. Descripción conjunta de varias variables



Francis Galton (1822-1911)
Científico y explorador británico. Inventor de la regresión. Viajó por África, donde hizo muchos descubrimientos geográficos y climatológicos, además de inventar el saco de dormir. Primo de Darwin, dedicó la segunda parte de su vida a probar la teoría de la evolución. Sus trabajos sobre las huellas dactilares condujeron a su uso para la identificación policial.

3.1 Distribuciones de frecuencias multivariantes

Uno de los objetivos del análisis estadístico es encontrar las relaciones que existen entre un grupo de variables. En este capítulo presentamos una introducción a los métodos para cuantificar estas relaciones. Supondremos inicialmente, para simplificar, que el conjunto de datos contiene los valores de dos variables (x, y), que se han medido conjuntamente en ciertos elementos de una población. Posteriormente, este análisis se generaliza para cualquier número de variables.

3.1.1 Distribución conjunta

Llamaremos distribución conjunta de frecuencias de dos variables (x, y) a una tabla que representa los valores observados de ambas variables y las frecuencias relativas de aparición de cada par de valores. Siempre conviene dar el número de elementos observados de manera que podamos calcular también inmediatamente las frecuencias absolutas si se desea. Cuando las variables son cualitativas, la tabla resultante se denomina tabla de contingencias, reservándose el nombre de distribución conjunta para variables numéricas. La construcción de buenas tablas de frecuencias no es inmediata, y el apéndice 3C de este capítulo presenta algunos principios generales.

La tabla 3.1 presenta una tabla de contingencia con las frecuencias relativas del resultado de observar el color de los ojos de 1.000 personas (variable hijo) y preguntarles por el color de los ojos de su madre. Se observa que

Tabla 3.1 Frecuencias relativas del color de ojos de 1.000 personas y de sus madres

	Madres		
Hijo	**Claros**	**Oscuros**	**Total**
Claros	0,25	0,08	0,33
Oscuros	0,12	0,55	0,67
TOTAL	0,37	0,63	

Tabla 3.2 Frecuencias relativas mensuales de asistencia al cine y al teatro para una muestra de 200 estudiantes universitarios

	Teatro			
Cine	**0**	**1**	**2**	**Total**
1	0,41	0,05	—	0,46
2	0,19	0,06	0,02	0,27
3	0,10	0,05	0,02	0,17
4	0,02	0,07	0,01	0,10
TOTAL	0,72	0,23	0,05	1

la combinación más frecuente es oscuros-oscuros, seguida de claros-claros. En los márgenes de la tabla se han sumado las frecuencias relativas por filas y por columnas. La tabla 3.2 presenta las frecuencias relativas de asistencia al cine y al teatro en un mes dado para una muestra de 200 estudiantes universitarios.

En estas dos tablas el interior de cada casilla (x_i, y_j) contiene la frecuencia relativa $fr(x_i, y_j)$ correspondiente a los dos valores que definen la casilla. Por tanto:

$$\sum_i \sum_j fr(x_i, y_j) = 1$$

Las frecuencias absolutas de las casillas se obtienen multiplicando el total de elementos por la frecuencia relativa. Por ejemplo, la frecuencia absoluta de la casilla (1,0) de la tabla 3.2 con frecuencia relativa 0,41 es $0{,}41 \times 200 = 82$ personas. Esta idea de representación conjunta puede extenderse para cualquier número de variables, aunque la representación gráfica no sea posible para más de tres.

Cuando las dos variables no toman valores repetidos, como suele ocurrir con variables continuas, la distribución conjunta se obtiene agrupando las dos variables en clases, como hicimos en el caso univariante, y calculando las frecuencias relativas de las casillas correspondientes. La tabla 3.3 presenta un ejemplo de distribución conjunta con datos agrupados.

Tabla 3.3 Frecuencias relativas del volumen de ventas y número de trabajadores para un grupo de 100 empresas pequeñas y medianas

Ventas	1-24	25-59	50-74	75-99	Total
1-100	0,28	0,07	0,01	0,00	0,36
101-200	0,10	0,15	0,06	0,02	0,33
201-300	0,04	0,10	0,08	0,09	0,31
TOTAL	0,42	0,32	0,15	0,11	

3.1.2 Distribuciones marginales

Se denomina distribución marginal de una variable a la obtenida al estudiar la variable aisladamente, con independencia del resto. El nombre de marginal proviene de que esta distribución se obtiene a partir de la distribución conjunta acumulando en los márgenes de la tabla la suma de las frecuencias relativas de las filas o columnas. En general, si llamamos $fr(x_i, y_j)$ a las fre-

cuencias relativas de la distribución conjunta, las frecuencias relativas que definen la distribución marginal de x se obtienen con:

$$fr(x_i) = \sum_j fr(x_i, y_j) \tag{3.1}$$

y análogamente:

$$fr(y_j) = \sum_i fr(x_i, y_j) \tag{3.2}$$

Las tablas 3.1, 3.2 y 3.3 presentan ejemplos de distribuciones marginales, que aparecen en los márgenes de las tablas. Por ejemplo, la distribución marginal del color de los ojos de las madres en la tabla 3.1 toma dos posibles valores, claros y oscuros, con frecuencias relativas 0,37 y 0,63. En la tabla 3.2 la distribución marginal de la variable número de asistencias al teatro toma los valores posibles 0, 1, 2 con frecuencias relativas 0,72, 0,23 y 0,05 respectivamente. En la tabla 3.3 las ventas de las empresas están en los intervalos (1-100), (101-200) y (201-300) con frecuencias relativas 0,36, 0,33 y 0,31.

3.1.3 Distribuciones condicionadas

La distribución condicionada de y para $x = x_i$ es la distribución univariante de la variable y que se obtiene considerando sólo los elementos que tienen para la variable x el valor x_i. Puede obtenerse de la distribución conjunta dividiendo las frecuencias relativas de la línea definida por $x = x_i$ por su suma. Llamando $fr(y_j|x_i)$ a las frecuencias relativas de esta distribución:

$$\boxed{fr(y_j|x_i) = \frac{fr(x_i, y_j)}{fr(x_i)}} \tag{3.3}$$

Con esta operación garantizamos que la suma de las frecuencias relativas para todos los valores de la variable y es uno, ya que, sumando para los valores de y:

$$\sum fr(y_j|x_i) = \frac{\sum fr(x_i, y_j)}{f(x_i)} = 1$$

Por ejemplo, llamando y a la variable asistencia al cine y x a la variable asistencia al teatro, de la tabla 3.2, la distribución de asistencia al cine para los estudiantes que no van nunca al teatro se presenta en la tabla 3.4.

Tabla 3.4 Distribución condicionada del número de asistencias al cine para los estudiantes que no han ido al teatro

Cine	0
1	0,41/0,72 = 0,57
2	0,19/0,72 = 0,26
3	0,10/0,72 = 0,14
4	0,02/0,72 = 0,03

En general la distribución condicionada de y para $x = x_i$ puede interpretarse como la distribución de la característica y en los elementos de la población que tienen como característica x el valor x_i. Se diferencia de la distribución marginal de y en que ésta tiene en cuenta la distribución de y en todos los elementos, con independencia del valor que en ellos tenga la característica x. De (3.1) y (3.3) se deduce que:

$$fr(y) = \sum_i fr(y|x_i)fr(x_i)$$

que indica que la frecuencia de la característica y en la población total puede obtenerse ponderando su frecuencia en las subpoblaciones definidas por distintos valores de x por el peso relativo de cada subpoblación en la población total.

La ecuación (3.3) establece que si conocemos las distribuciones condicionadas de y dada x y la distribución marginal de x, podemos calcular la distribución conjunta mediante:

$$fr(x_i, y_j) = fr(y_j|x_i)fr(x_i)$$

Por tanto, si conocemos todas las distribuciones condicionadas y las marginales para cada variable podemos calcular la distribución conjunta.

Ejemplo 3.1

Calcular la distribución del número de trabajadores condicionada a unas ventas en el intervalo (101-200). Indicar la distribución con frecuencias relativas y absolutas. La fila correspondiente de la tabla es

101-200 0,10 0,15 0,06 0,02 0,33

y para calcular la distribución dividimos cada frecuencia relativa por la suma de todas las casillas, 0,33. El resultado es, redondeando:

	Trabajadores			
Ventas	**1-24**	**25-59**	**50-74**	**75-99**
101-200	0,10/0,33 = 0,30	0,15/0,33 = 0,45	0,06/0,33 = 0,18	0,02/0,33 = 7

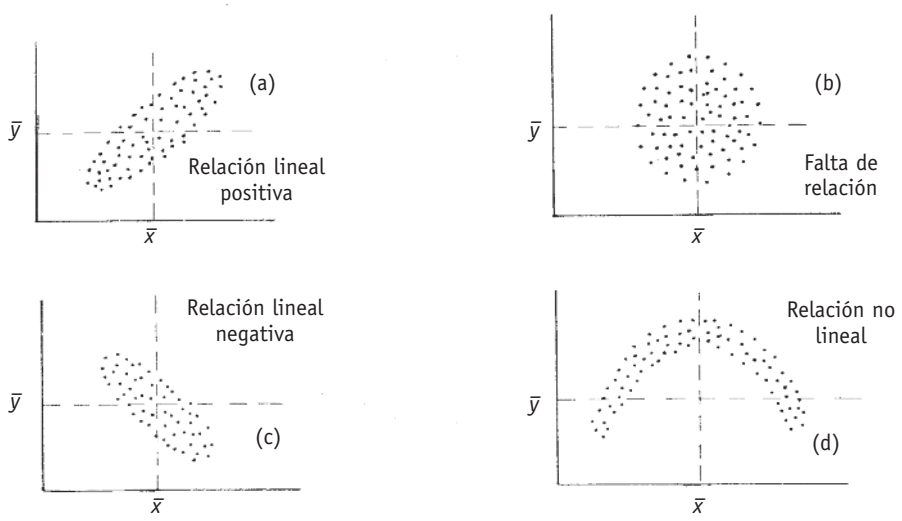
Para obtener las frecuencias absolutas multiplicamos estas frecuencias relativas por el total de empresas que estamos considerando. Como de las 200 empresas el 33% tiene ventas entre 101 y 200, serán $200 \times 0,33 = 66$ empresas. Las frecuencias absolutas serán, multiplicando por 66 las frecuencias relativas y redondeando:

	Trabajadores			
Ventas	**1-24**	**25-59**	**50-74**	**75-99**
101-200	20	30	12	46

3.1.4 Representaciones gráficas

La representación gráfica más útil de dos variables continuas sin agrupar es el diagrama de dispersión, que se obtiene representando cada observación bidimensional $(x_i y_i)$ como un punto en el plano cartesiano. Este diagrama es especialmente útil para indicar si existe o no relación entre las variables. La figura 3.1 presenta algunos ejemplos.

Figura 3.1 Distintos tipos de relación entre las variables



Para datos agrupados podríamos construir diagramas de barras o histogramas en dos dimensiones. Estas representaciones, que pueden hacerse con un ordenador, se utilizan poco.

Ejemplo 3.2

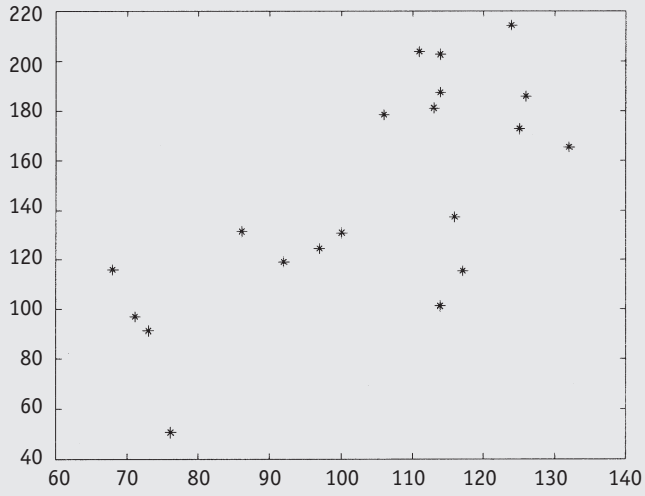
La tabla adjunta indica el precio de venta en miles de euros y la superficie de un conjunto de 20 viviendas. Estudiar la relación entre ambas variables con un diagrama de dispersión

$\mathbf{m^-}$	106	73	114	132	86	117	125	68	71	111	92
Euros	178	91	188	165	132	115	173	116	97	204	119

$\mathbf{m^-}$	114	116	114	126	113	124	76	100	97
Euros	101	137	203	186	181	214	50	131	124

La figura 3.2 presenta el gráfico de dispersión.

Figura 3.2 Precio de la vivienda en miles de euros y superficie en m^2



Se observa una relación entre las variables en el sentido de que al aumentar la superficie aumenta, en promedio, el precio de la vivienda.

3.2 Medidas de dependencia lineal

3.2.1 Covarianza

En el estudio conjunto de variables continuas interesa disponer de una medida descriptiva de la relación lineal entre cada par de variables. La medida más utilizada es la covarianza, definida por:

$$Cov(x, y) = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{n} \tag{3.4}$$

donde el sumatorio está extendido a las n parejas de valores (x, y). Una expresión equivalente de la covarianza es

$$Cov(x, y) = \frac{\sum x_i y_i}{n} - \overline{x}\,\overline{y}$$

Para datos agrupados en clases la fórmula anterior se reduce a

$$Cov(x, y) = \sum_i \sum_j (x_i - \overline{x})(y_j - \overline{y})\, fr(x_i, y_i) \tag{3.5}$$

y ahora el sumatorio está extendido a todas las clases.

La covarianza fue introducida por K. Pearson para medir la relación lineal entre x e y. Para ilustrarlo, consideremos los diagramas de dispersión de la figura 3.1. Vamos a comprobar que cuando x e y varían conjuntamente de forma lineal, como indican los casos 3.1(a) y (c), la covarianza será alta en valor absoluto, aunque positiva en el caso (a) y negativa en el caso (b). Consideremos los cuadrantes definidos por los ejes que pasan por el punto medio de los datos $(\overline{x}, \overline{y})$. En el caso (a) la mayoría de las desviaciones $x_i - \overline{x}$ e $y_i - \overline{y}$ estarán en el primer y tercer cuadrantes. Como ambas desviaciones tienen en ambos cuadrantes el mismo signo, su producto será positivo y la covarianza será positiva y alta. El signo positivo de la covarianza indica que cuando una variable está por encima de la media, es esperable que la otra también lo esté, como vemos en el gráfico. En el caso (c) la mayoría de las desviaciones $x_i - \overline{x}$ e $y_i - \overline{y}$ están en el segundo y cuarto cuadrantes. En estos cuadrantes las desviaciones tienen signos opuestos, y la covarianza será alta en magnitud, pero negativa.

Por el contrario, cuando no existe relación —caso (b)— o existe relación no lineal —caso (d)—, la covarianza será pequeña al estar los puntos repartidos por los cuatro cuadrantes. Como en dos el producto es positivo y en otros dos negativo, los términos se cancelarán aproximadamente y la covarianza será baja. Observemos que esto ocurre tanto en el caso (b),

donde no existe relación, como en el (d), donde existe una clara relación no lineal. Esto es así porque la covarianza se inventó para medir relaciones lineales.

3.2.2 Correlación

El inconveniente de la covarianza como medida de asociación es su dependencia de las unidades de medida de las variables: supongamos que la covarianza entre la estatura, medida en centímetros, y el peso en gramos en unos datos es 200; si expresamos la estatura en metros, los valores de las estaturas quedan divididos por 100, y si ahora expresamos los pesos en kilogramos, dividiremos los pesos por 1.000. En consecuencia, la covarianza entre el peso y la altura en las nuevas unidades será, ahora, 0,002.

Para construir una medida adimensional de la relación lineal entre dos variables tendremos que dividir la covarianza por un término que tenga sus mismas dimensiones. Como la covarianza va en el producto de las unidades de las variables, Galton propuso definir el coeficiente de correlación entre dos variables por:

$$r = \frac{Cov(x, y)}{s_x s_y} \tag{3.6}$$

donde s_x y s_y son las desviaciones típicas de x y de y. El lector debe comprobar (véanse los ejercicios 3.1 y 3.2 al final del capítulo) que:

1. El coeficiente de correlación tiene el mismo signo que la covarianza.
2. El coeficiente de correlación es adimensional: su valor no varía si multiplicamos x por k_1 e y por k_2, siendo k_1 y k_2 números no nulos del mismo signo.
3. Si existe una relación lineal exacta entre ambas variables, lo que supone que todos los puntos deben estar en una línea recta, que podemos escribir como $y = a + bx$, el coeficiente de correlación es igual a 1 (si $b > 0$) o -1 (si $b < 0$).
4. Si no existe una relación lineal exacta (los puntos no están sobre una recta), $-1 < r < 1$.

Es importante recordar que el coeficiente de correlación es una medida *resumen* de la estructura de un diagrama de dispersión y que, en consecuencia, *siempre* conviene dibujar este diagrama que contiene toda la infor-

mación. Por ejemplo los diagramas (b) y (d) de la figura 3.1 conducen ambos a una correlación muy próxima a cero y, sin embargo, corresponden a situaciones muy distintas.

Ejemplo 3.3

Calcular la covarianza y el coeficiente de correlación para los datos del ejemplo 3.2 de superficies y precios.

La media de las superficies (x) es 103,75, y la media de los precios (y), 145,214 miles de euros. La covarianza será

$$Cov(x, y) = \frac{(106 - 103{,}75)(178 - 145{,}214) + \ldots + (97 - 103{,}75)(124 - 145{,}214)}{20} = 414{,}8$$

mientras que las varianzas de las variables son
$Var(x) = 259{,}2$ y $Var(y) = 1316$. El coeficiente de correlación es

$$r = \frac{414{,}8}{\sqrt{259{,}2}\sqrt{1316}} = 0{,}71$$

3.3 Recta de regresión

Cuando dos variables están relacionadas de forma lineal, los puntos tienden a agruparse en el diagrama de dispersión alrededor de una recta. Un procedimiento natural de expresar esta relación es mediante la recta que describe su evolución conjunta. De la misma forma que describimos una variable por la media y la dispersión, podemos describir la relación entre dos variables por una recta y la dispersión de los puntos con relación a esa recta.

La media de una variable minimiza las diferencias entre los datos y la media, que son en promedio cero. Podemos aplicar la misma idea para construir la recta media. Para simplificar, supongamos que estamos interesados en minimizar los errores de la variable y cuando conocemos el valor de x. Éste es el enfoque natural si deseamos prever y dado x. Entonces la recta será de la forma

$$h(x) = a + bx$$

donde a es la ordenada en el origen [valor de $h(x)$ cuando $x = 0$] y b será la pendiente, que es el incremento de $h(x)$ si x aumenta una unidad.

Si decidimos medir las distancias en el sentido vertical, la recta resultante se denomina *recta de regresión*.

Los coeficientes a y b se determinan minimizando las distancias verticales entre los puntos observados, y_i, y las ordenadas previstas por la recta para dichos puntos, $a + bx_i$. El criterio será minimizar:

$$\sum (y_i - a - bx_i)^2 \tag{3.7}$$

donde las desviaciones se han tomado al cuadrado para prescindir de su signo. Derivando respecto a ambos coeficientes e igualando a cero, resultan las ecuaciones:

$$2\sum (y_i - a - bx_i)(-1) = 0$$

$$2\sum (y_i - a - bx_i)(-x_i) = 0$$

Dividiendo por n, número de parejas (x_i, y_i) observadas, estas ecuaciones pueden escribirse:

$$\bar{y} = a + b\bar{x} \tag{3.8}$$

$$\frac{\sum x_i y_i}{n} = a\bar{x} + b\frac{\sum x_i^2}{n} \tag{3.9}$$

La primera ecuación indica que la recta debe pasar por el centro de la nube de puntos $(\bar{x}, \bar{y})$. Eliminando a de la segunda ecuación restando (3.8) multiplicada por $\bar{x}$ de (3.9) se obtiene

$$\frac{\sum x_i y_i}{n} - \bar{x}\bar{y} = b\left(\frac{\sum x_i^2}{n} - \bar{x}^2\right)$$

El primer miembro es la covarianza entre ambas variables, y el segundo, b veces la varianza de x, por lo que la pendiente de la recta es:

$$\boxed{b = \frac{Cov(x, y)}{s_x^2}} \tag{3.10}$$

exprexión que indica que la pendiente de la recta es la covarianza estandarizada para que tenga unidades de y / x como corresponde a la pendiente. Observemos que la estandarización se obtiene con la desvición típica, de manera similar a la estandarización de la covarianza para obtener el

coeficiente de correlación. Sustituyendo en la ecuación de la recta $a = \bar{y} - b\bar{x}$ y b por su expresión (3.10), la recta se calcula como:

$$h(x) = \bar{y} + \frac{Cov(x, y)}{s_x^2}(x - \bar{x}) \tag{3.11}$$

Esta recta se denomina recta de regresión en honor a Galton, que la obtuvo por primera vez tomando como x las estaturas de padres e y las estaturas de los hijos. Galton obtuvo que la pendiente de la relación es menor que la unidad, lo que implica que cuando la estatura de un padre es mucho mayor que la media, la estatura esperada de sus descendientes será también mayor que la media, pero menor que la del padre. Este fenómeno, de gran importancia en biología, se conoce como regresión a la media.

Podemos construir una medida de variabilidad de los datos respecto a la recta de regresión igual que hicimos con las desviaciones típicas promediando las desviaciones verticales al cuadrado entre cada punto y la ordenada correspondiente a la recta. Llamaremos *desviación típica residual* a:

$$s_R = \sqrt{\frac{\sum[y_i - h(x)]^2}{n}} \tag{3.12}$$

donde $h(x)$ es la recta de regresión dada por (3.11) y $(y_i x_i)$ son las coordenadas de cada punto. La desviación típica residual mide, en consecuencia, la desviación vertical promedio entre los puntos y la recta de regresión.

Observemos que si no hay relación entre x e y y la covarianza es nula, $b = 0$ y la recta se reduce a $h(x) = a = \bar{y}$. En consecuencia, la desviación típica residual se convierte en la desviación típica de la variable y.

Ejemplo 3.4

Los pesos (en kg) y estaturas (cm) de una muestra de 10 estudiantes universitarios son:

(x) Peso	82	75	70	68	44	63	80	79	54	54
(y) Estatura	185	185	180	178	159	170	190	172	162	165

Calcular la covarianza, el coeficiente de correlación, la recta de regresión de la estatura en función del peso y la desviación típica residual:

$$\bar{y} = 174{,}6; \quad s_y = 10{,}08; \quad \bar{x} = 66; \quad s_x = 11{,}62$$

Calcularemos la covarianza utilizando que:

$$Cov(x, y) = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n} = \frac{\sum x_i y_i}{n} - \bar{x}\bar{y}$$

Por tanto:

$$\sum x_i y_i = 116.353; \qquad Cov(x, y) = 11635.3 - (174.6)(66) = 111.7$$

$$r = \frac{111.7}{(19.08)(11.62)} = 0.95$$

La pendiente de la recta es:

$$b = \frac{111.7}{(11.62)^2} = 0.83$$

y la ordenada en el origen:

$$a = \bar{y} - b\bar{x} = 174.6 - 0.83 \cdot 66 = 119.82$$

luego la recta de regresión es:

$$h(x) = 119.82 + 0{,}83x$$

y la desviación típica residual se calcula como:

$$\sum[y_i - h(x_i)]^2 = (185 - 119.82 - 0.83 \cdot 82)^2 + \ldots + (165 - 119.82 - 0.83 \cdot 54)^2 =$$

$$= 8.92 + 8.54 + 4.33 + \ldots + 6.97 + 0.13 = 92.19$$

$$s_R = \sqrt{\frac{92.19}{10}} = 3.03$$

Este resultado indica que la desviación promedio entre las estaturas observadas y las previstas con la recta de regresión es de 3 cm.

3.3.1 Correlación y regresión

La covarianza, el coeficiente de correlación y la pendiente de la recta que describe la nube de puntos son tres formas estrechamente relacionadas de expresar la dependencia lineal. El coeficiente de correlación es adimensio-

nal (no cambia al expresar las variables en otras unidades), mientras que la covarianza tiene unidades de (xy) y la pendiente de la recta de $(y|x)$.

El coeficiente de correlación es simétrico en ambas variables, ya que mide la relación. Sin embargo, la recta de regresión no lo es porque se construye suponiendo que el valor de una variable es conocido (el de la x) y que queremos prever la otra (y).

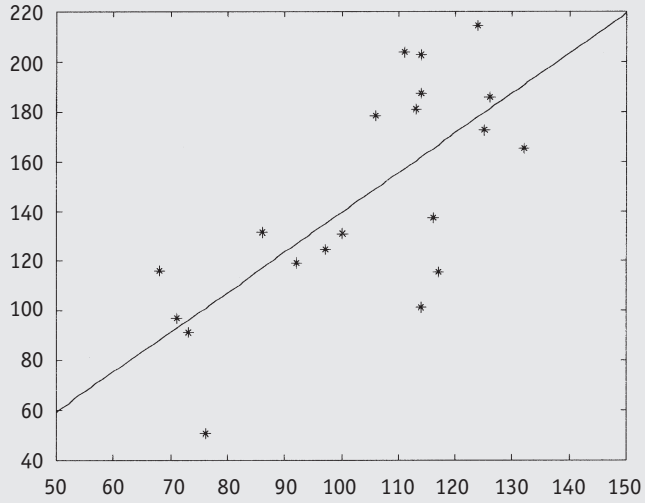
Ejemplo 3.5

Calcularemos la recta de regresión para prever el precio de un piso en euros dada su superficie con los datos del ejercicio 3.3.

La pendiente será $b = cov(x, y)/var(x) = 414{,}8/259{,}2 = 1{,}6$ miles de euros por m^2. La ordenada en el origen $a = 145{,}214 - 1{,}6 \times 103{,}75 = -20{,}78$.

La ecuación es por lo tanto Precio = – 20,78 + 1,6 m^2. Por ejemplo, el precio previsto por esta ecuación para un piso de 80 m^2 es Precio = – 20,78 + + 1,6 × 80 = 107,21 miles de euros. La figura 3.3 indica el gráfico de los puntos y la recta de regresión.

Figura 3.3 Recta de regresión entre el precio de un piso y su superficie



3.4 Vector de medias

En el estudio de variables cuantitativas k-dimensionales, las k observaciones asociadas a un individuo pueden considerarse como un vector $\mathbf{X}$, cuyos

componentes son los valores que en él toma cada variable. El conjunto de datos se representa por la secuencia de vectores $\mathbf{X}_1$, ..., $\mathbf{X}_n$. Llamaremos vector de medias de la variable *k*-dimensional al vector $\bar{\mathbf{X}}$ de dimensión *k* cuyos componentes son las medias aritméticas de cada variable.

Por ejemplo, para una variable tridimensional:

$$\mathbf{X}_j = \begin{bmatrix} x_i \\ y_i \\ x_i \end{bmatrix}$$

tendremos:

$$\bar{\mathbf{X}} = \begin{bmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \end{bmatrix} = \frac{1}{n} \begin{bmatrix} \sum x_i \\ \sum y_i \\ \sum z_i \end{bmatrix} = \frac{1}{n} \sum \mathbf{X}_i$$

ya que los vectores se suman sumando sus componentes. Por tanto, en general:

$$\bar{\mathbf{X}} = \frac{1}{n} \sum \mathbf{X}_i \tag{3.13}$$

3.5 Matriz de varianzas y covarianzas

Llamaremos matriz de varianzas y covarianzas —o simplemente matriz de covarianzas— a la matriz cuadrada simétrica que tiene en la diagonal principal las varianzas de las observaciones y fuera de ellas las covarianzas entre variables. Por ejemplo, en el caso bidimensional:

$$\mathrm{M} = \begin{bmatrix} s_x^2 & cov(x, y) \\ cov(x, y) & s_x^2 \end{bmatrix}$$

Esta matriz será siempre simétrica, ya que $cov(x, y) = cov(y, x)$. En el caso de una variable *k*-dimensional, llamando s_i^2 a la varianza del componente *i*, y s_{ij} a la covarianza entre las variables *i* y *j*, la matriz de varianzas y covarianzas es:

$$\mathrm{M} = \begin{bmatrix} s_1^2 & s_{12} & \cdots & s_{1k} \\ s_{21} & s_2^2 & \cdots & s_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ s_{\mathrm{k1}} & s_{k2} & \cdots & s_k^2 \end{bmatrix}$$

Utilizando la notación vectorial, la matriz de varianzas y covarianzas se calcula, conocido el vector de medias $\bar{\mathbf{X}}$, por:

$$M = \frac{1}{n}\sum (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}}) \tag{3.14}$$

donde el sumatorio va extendido al conjunto de todos los elementos estudiados. En efecto, para una variable tridimensional:

$$M = \frac{1}{n}\sum \begin{bmatrix} x_i - \bar{x} \\ y_i - \bar{y} \\ z_i - \bar{z} \end{bmatrix} [x_i - \bar{x} \quad y_i - \bar{y} \quad z_i - \bar{z}]$$

$$M = \frac{1}{n}\sum \begin{bmatrix} (x_i - \bar{x})^2 & (x_i - \bar{x})(y_i - \bar{y}) & (x_i - \bar{x})(z_i - \bar{z}) \\ (y_i - \bar{y})(x_i - \bar{x}) & (y_i - \bar{y})^2 & (y_i - \bar{y})(z_i - \bar{z}) \\ (z_i - \bar{z})(x_i - \bar{x}) & (z_i - \bar{z})(y_i - \bar{y}) & (z_i - \bar{z})^2 \end{bmatrix}$$

Observemos que las fórmulas del vector de medias y de la matriz de varianzas y covarianzas son análogas a las de la media y varianza para escalares sustituyendo:

1. Escalares por vectores.
2. Cuadrados de escalares por producto por el transpuesto para vectores.

Ejemplo 3.6

El vector de medias en la distribución conjunta de precios en miles de euros y superficie de m² con los datos del ejemplo 3 es

$$\begin{bmatrix} 103{,}75 \\ 145{,}214 \end{bmatrix}$$

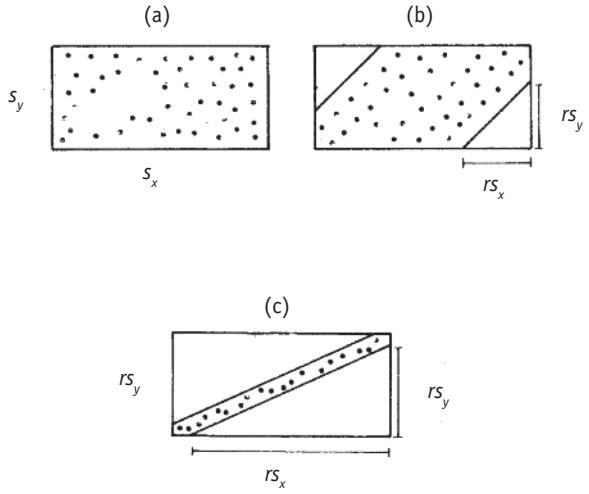
y la matriz de varianzas y covarianzas

$$\begin{bmatrix} 259{,}2 & 414{,}8 \\ 414{,}8 & 1316 \end{bmatrix}$$

3.5.1 Varianza efectiva

Una medida global escalar de la variabilidad conjunta de *k*-variables es la varianza efectiva, que es la raíz de orden *k* del determinante de la matriz de varianzas y covarianzas. Su raíz cuadrada se denomina desviación típica efectiva, y tiene las propiedades siguientes:

Figura 3.4 La varianza promedio como una medida de dispersión conjunta



a) Está bien definida, ya que el determinante de la matriz de varianzas y covarianzas es siempre positivo, como demostraremos en el apéndice 2D.
b) Es una medida de la variabilidad promedio del conjunto de datos.

Para aclarar estas ideas supongamos el caso $k = 2$. Entonces, utilizando la definición del coeficiente de correlación, M puede escribirse:

$$\mathrm{M} = \begin{bmatrix} s_x^2 & rs_x s_y \\ rs_x s_y & s_y^2 \end{bmatrix}$$

y la varianza efectiva es:

$$VE = |\mathrm{M}|^{1/2} = \sqrt{s_x^2 s_y^2 (1 - r^2)} \tag{3.15}$$

Si las variables son independientes, la mayoría de sus valores estarán dentro de un rectángulo de lados $6s_x$, $6s_y$, ya que, por el teorema de Tchebychev, entre la media y 3 desviaciones típicas debe estar aproximadamente el 90% de los datos. En consecuencia, el área ocupada por ambas variables es directamente proporcional al producto de las desviaciones típicas [figura 3.4(a)].

Si las variables están relacionadas linealmente, el coeficiente de correlación será distinto de cero. Supongamos que sea positivo, como en la figura 3.4(b). Entonces, la mayoría de los puntos tienden a situarse en una

franja como la indicada, y habrá una reducción del área tanto mayor cuanto mayor sea r [figuras 3.4(b) y (c)]. En el límite, si $r = 1$, todos los puntos están en una línea, hay una relación lineal exacta entre las variables y el área ocupada es cero. La fórmula (3.15) describe esta contracción del área ocupada por los puntos al aumentar el coeficiente de correlación.

La desviación típica efectiva será

$$DM = \sqrt{s_x s_y \sqrt{1 - r^2}}$$

Si las variables están incorreladas esta medida es la media geométrica de las desviaciones típicas. Cuando las variables están incorreladas, la medida incluye un término que tiene en cuenta la dependencia lineal entre las variables.

Ejemplo 3.7

La tabla proporciona tres indicadores económicos para los países de la OCDE (los datos siguen el orden alfabético).

x_+ Tasa real de crecimiento del PNB media 72/82	x_- Tasa de desempleo en 1986	$x_\times$ Tasa de incremento de índices de precios 1985
2,0	8,00	2,2
2,8	7,50	7,6
2,6	4,50	2,9
2,2	13,50	4,6
2,8	9,50	4,1
1,8	8,75	3,9
2,6	22,50	7,4
2,2	7,25	3,2
3,1	6,25	5,1
2,7	10,75	5,3
3,1	9,00	20,1
1,9	14,00	2,3
4,0	16,75	5,5
3,4	1,00	32,7
2,6	10,50	9,1
4,3	2,75	1,7
1,7	1,75	3,2
4,0	2,25	5,8

x_+ Tasa real de crecimiento del PNB media 72/82	x_- Tasa de desempleo en 1986	$x_\times$ Tasa de incremento de índices de precios 1985
1,6	4,75	16,3
3,8	11,50	15,9
1,5	11,50	5,9
1,6	3,00	6,7
0,6	0,50	3,4
5,1	13,50	40,5

La tabla siguiente muestra algunas medidas características de estas tres variables:

	x_+	x_-	$x_\times$
Media	2,67	8,38	8,97
Mediana	2,60	8,00	5,30
D. típica	1,05	5,43	9,77
CV	0,39	0,64	1,09
C. asimetría	0,36	0,53	1,98
C. curtosis	2,64	2,86	6,10

Se observa que la mayor homogeneidad (menor CV y coeficiente de curtosis próximo a tres) entre los países de la OCDE aparece con la variable crecimiento seguida del desempleo. Los datos de inflación muestran alta heterogeneidad y el alto coeficiente de curtosis hace sospechar la presencia de datos atípicos. Esto se confirma en los histogramas de la figura 3.5, donde las variables se representan en su orden.

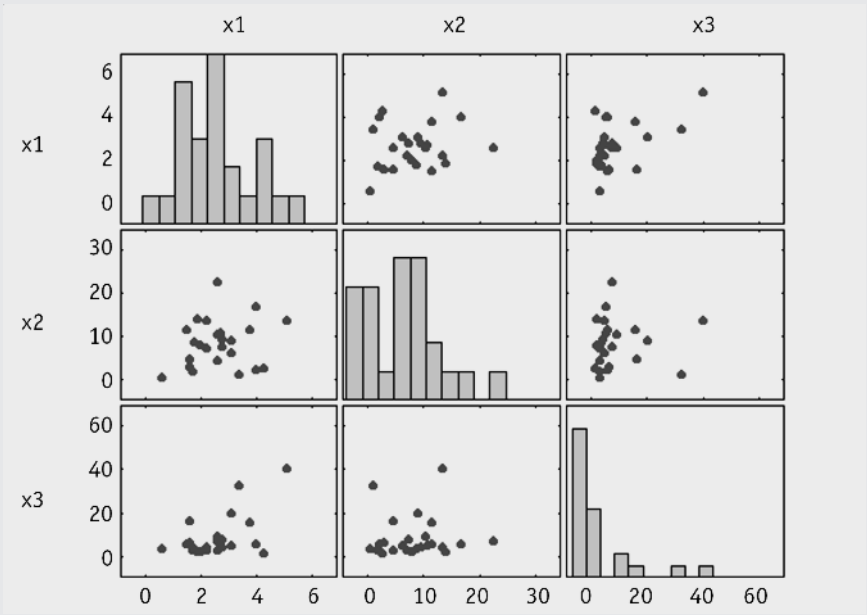
La relación entre estas tres variables vendrá dada por la matriz de varianzas y covarianzas. Para calcularlas, como:

$$\sum(x_i - \bar{x})(y_i - \bar{y}) = \sum(x_i - \bar{x})y_i - \bar{y}\ \sum(x_i - \bar{x}) = \sum(x_i - \bar{x})y_i$$

sólo es necesario obtener las desviaciones a la media por una variable. Por tanto:

$$Cov(x_1, x_2) = \frac{1}{24}[(2{,}0 - 2{,}67)8 + (2{,}8 - 2{,}67)7{,}5 + \ldots + (5{,}1 - 2{,}67)13{,}5] = 1{,}11$$

Figura 3.5 Representación conjunta de los histogramas y los gráficos de dispersión para las variables



Procediendo análogamente con las restantes variables se obtiene la matriz de varianzas y covarianzas

$$M = \begin{bmatrix} 1{,}10 & 1{,}11 & 5{,}12 \\ 1{,}11 & 29{,}50 & 2{,}12 \\ 5{,}12 & 2{,}12 & 95{,}58 \end{bmatrix}$$

y los correspondientes coeficientes de correlación son:

$$r_{12} = \frac{1{,}11}{1{,}05 \cdot 5{,}43} = 0{,}19$$

$$r_{13} = \frac{5{,}12}{1{,}05 \cdot 9{,}77} = 0{,}50$$

$$r_{23} = \frac{2{,}12}{5{,}43 \cdot 9{,}77} = 0{,}04$$

que muestran que la única correlación apreciable se da entre las variables 1 y 3. El gráfico siguiente amplía la información de la matriz de varianzas y covarianzas indicando en la diagonal los histogramas de cada variable y fuera de la diagonal los diagramas de dispersión entre pares de variables.

Ejercicios 3

3.1. Demostrar que al multiplicar x por k_1 e y por k_2 el coeficiente de correlación entre ambas no varía (k_1 y k_2 deben tener el mismo signo).

3.2. Demostrar que si entre dos variables existe una relación exacta $y = a + bx$, con $b > 0$, el coeficiente de correlación es uno.

3.3. Demostrar que el coeficiente de correlación es siempre en valor absoluto menor que uno.

3.4. Calcule la covarianza y el coeficiente de correlación de los datos siguientes:

x	2	1	2	1	4
y	8	9	8	5	10

3.5. Calcule la recta de regresión para los datos del ejercicio anterior.

3.6. Intercambie los valores de x y de y en el ejercicio 3.4 y calcule la nueva recta. Compruebe que el producto de esta pendiente y la encontrada en 3.5 es el coeficiente de correlación al cuadrado. ¿Ocurrirá esto siempre?

3.7. Demuestre que cuando la variable x es un atributo que toma únicamente los valores – 1 (n_1 veces) y + 1 (n_2 veces), la covarianza con n datos ($n = n_1 + n_2$) entre x y una variable continua y es $2n_1n_2\,(m_1 - m_2)/n^2$, siendo m_1, m_2 las medias de y para ambos valores de x.

3.8. Obtenga datos de la altura y el peso de 20 personas. Haga un gráfico entre ambas variables y calcule el coeficiente de correlación entre ellas.

3.9. Obtenga de la web del Instituto Nacional de Estadística, http://www.ine.es, los datos de variables demográficas de las comunidades autónomas españolas y calcule su vector de medias y la matriz de varianzas y covarianzas. Estudie la relación entre las variables demográficas mediante diagramas de dispersión.

3.10. Construya un diagrama de dispersión y calcule la recta de regresión para prever la velocidad de una galaxia en función de la distancia con los datos siguientes, donde aparecen la distancia de ciertas galaxias en millones de años luz y la velocidad en miles de millas por segundo:

Distancia	22	68	108	137	255	315	390	405	685	700	1100
Velocidad	75	2,4	3,2	4,7	9,3	12,0	13,4	14,4	24,5	26,0	38

3.11. La tabla adjunta indica la proporción de su renta que una muestra de hogares se gasta en alimentación. Ajustar una recta de regresión para explicar la proporción

del gasto en alimentación en función de la renta por persona con los datos siguientes; la renta se da en miles de euros al año por persona.

Porporción	22	24	25	28	30	33	37	40	42	42
Renta	30	27	22	23	19	20	15	14	11	12

3.6 Resumen del capítulo y consejos de cálculo

En este capítulo hemos estudiado cómo describir la dependencia de un conjunto de variables. Toda la información sobre su dependencia está incluida en su distribución conjunta. Para varias variables es normalmente más simple estudiar las distribuciones condicionadas de una variable con las restantes. La medida principal de dependencia lineal entre dos variables continuas es el coeficiente de correlación que se obtiene estandarizando la covarianza. La existencia de dependencia lineal entre dos variables implica que una variable puede preverse mejor conociendo el valor de la otra que sin esta información. La previsión de una variable dada la otra se efectúa con la recta de regresión. La variabilidad de un conjunto de variables se mide por la matriz de varianzas y covarianzas.

El cuadro 3.1 resume las fórmulas principales introducidas en este capítulo.

Los métodos presentados en este capítulo hacen imprescindible el uso del ordenador. Excel permite realizar gráficos de dispersión entre dos variables y estimar la ecuación de regresión mediante el comando *Estimación lineal*, incluido entre las funciones estadísticas. Todos los programas estadísticos permiten calcular la recta de regresión; basta especificar las variables que se quieren relacionar y la matriz de varianzas y covarianzas entre un conjunto de variables definidas. En Statgraphics hay que ir a la barra desplegable del menú de comandos y elegir *Relate* y después escoger *Simple regression*. En Minitab hay que elegir *Regression*. En este programa la matriz de varianzas y covarianzas se encuentra en el apartado de *Multivariable*.

3.7 Lecturas recomendadas

Las técnicas de regresión se estudian con detalle en el segundo tomo del libro. Un estudio descriptivo más amplio se encuentra en muchos de los textos indicados en el apartado de análisis de datos. Ehrenberg (1986) y Mosteller y Tukey (1977) son especialmente recomendables. El estudio descriptivo de vectores de datos se aborda en los primeros capítulos de los textos de análisis estadístico multivariante. Estas técnicas se conocen también con el nom-

Cuadro 3.1 Fórmulas principales del capítulo 3

Relación entre frecuencias conjuntas, marginales y condicionadas	$fr(x_i y_j) = fr(x_i \mid y_i) fr(y_i)$
Covarianza	$Cov(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n}$
Coeficiente de correlación	$r = \frac{Cov(x, y)^n}{s_x s_y}$
Recta de regresión	$h(x) = a + bx$
Coeficiente de regresión	$b = \frac{Cov(x, y)}{s_x^2}$
Ordenada en el origen	$a = \bar{y} - b\bar{x}$
Desviación típica residual	$s_R = \sqrt{\frac{\sum (y_i - a - bx_i)^2}{n}}$

bre de minería de datos. En español véase Lebart *et al.* (1985), Cuadras (1996), y Peña (2002); en inglés Barnett (1981), Flury (1997) y Krzanowski (2000).

Apéndice 3A: Números índice

Supongamos que se desea estudiar la evolución de una variable x_t a lo largo del tiempo. Un procedimiento es comparar sus valores con un valor inicial x_0 que tomaremos como origen. Por ejemplo, la evolución del precio de un bien puede describirse por el cociente p_t/p_0, donde p_t es el precio actual y p_0 el precio en un año base que tomaremos como origen.

Cuando interese describir la evolución conjunta de un grupo de variables $(x_{1t}, ..., x_{kt})$ a lo largo del tiempo, podemos acudir a la misma idea: tomar sus valores en un período concreto $(x_{10}, ..., x_{k0})$ como referencia y calcular un índice ponderado:

$$I_t = \sum a_i \frac{x_{it}}{x_{i0}}$$

donde los coeficientes a_i son positivos y suman uno y se obtienen en cada caso teniendo en cuenta la importancia relativa de la variable *i* en el conjunto a estudiar.

En los índices que miden la evolución de un conjunto de precios determinados, los precios se ponderan por su importancia económica. Por ejemplo, el índice de precios al consumo pondera los precios de distintos bienes por la proporción del gasto que supone cada bien en un presupuesto familiar medio. Estas ponderaciones pueden calcularse en el momento origen de la comparación (método de Laspeyres) o en el momento *t* que se compara (método de Paasche). Llamando q_{i0} a las cantidades adquiridas del bien *i* en el período origen, el índice de Laspeyres se calcula ponderando el precio de cada bien, p_{it}, por la proporción del gasto en dicho bien ($q_{i0} \cdot p_{i0}$) con relación al gasto total en el período origen ($\sum q_{i0} p_{i0}$). La fórmula resultante es:

$$I_t = \sum \frac{p_{it}}{p_{i0}} \left(\frac{q_{i0} p_{i0}}{\sum q_{i0} p_{i0}} \right) = \frac{\sum p_{it} q_{i0}}{\sum p_{i0} q_{i0}}$$

Otra alternativa es calcular las ponderaciones tomando las cantidades consumidas de cada bien en el período estudiado *t*, con lo que se obtiene el índice de Paasche:

$$I_t = \sum \frac{p_{it}}{p_{i0}} \left(\frac{q_{it} p_{i0}}{\sum q_{it} p_{i0}} \right) = \frac{\sum p_{it} q_{it}}{\sum p_{it} q_{i0}}$$

Apéndice 3B: Análisis descriptivo de series

La serie del gráfico 2.6 muestra una tendencia creciente, mientras que la de la figura 2.7 muestra un comportamiento oscilante, con mayor amplitud al final, alrededor del valor 51%. Diremos que la primera serie tiene tendencia y la segunda no. Una forma de describir la tendencia de una serie es ajustar a los datos una recta:

$$y_t = a + bt$$

donde la variable *t* representa el tiempo.

Si tomamos el criterio de minimizar las distancias entre los puntos observados y los valores de la recta en sentido vertical, las fórmulas de la recta de regresión permiten obtener los coeficientes *a* y *b*. Los cálculos se simplifican si la variable *t* (que hace el papel de *x* en estas fórmulas) tuviese media cero, lo que puede conseguirse definiendo *t* adecuadamente. Suponiendo que existen $n = 2k + 1$ datos (períodos observados), podemos definir *t* como:

$$t = (-k, -k + 1, ..., -1, 0, 1, ..., k - 1, k)$$

Entonces, como $\overline{t} = 0$, (3.8) se reduce a:

$$a = \sum_{-k}^{+k} \frac{y_t}{n} = \overline{y}$$

y la pendiente se estimará por:

$$b = \frac{\sum y_t t}{\sum t^2}$$

Puede comprobarse fácilmente que el coeficiente b así estimado es un promedio ponderado de los incrementos parciales. Por ejemplo, con 5 datos ($n = 2$) ($y_{-2}, y_{-1}, y_0, y_1, y_2$):

$$\sum t^2 = (-2)^2 + (-1)^2 + (0)^2 + (1)^2 + (2)^2 = 10$$

$$\sum t y_t = -2y_{-2} - y_{-1} + y_1 + 2y_2 = 2(y_{-1} - y_{-2}) + 3(y_0 - y_{-1}) + 3(y_1 - y_0) + 2(y_2 - y_1)$$

y llamando $b(i)$ al incremento en el período i:

$$b = 0{,}2\ b(-1) + 0{,}3\ b(0) + 0{,}3\ b(1) + 0{,}2\ b(2)$$

que indica que el crecimiento promedio en el período —medido por b, la pendiente de la recta— es una media ponderada de los crecimientos observados en cada uno de los períodos observados. La ponderación es simétrica respecto al período central, que tiene el peso máximo.

Este resultado sugiere que este procedimiento descriptivo no va a ser, en general, útil para la predicción, ya que los valores centrales reciben el máximo peso, mientras que los incrementos más recientes reciben el peso menor y análogo a los incrementos más alejados en el tiempo. Intuitivamente, un buen método de predicción debería dar más peso a los datos recientes que a los muy alejados. Esta intuición es acertada, y el lector interesado en la predicción de series temporales puede consultar este aspecto en la literatura especializada.

Apéndice 3C: La presentación de datos en tablas

La tabla 3C.1 está tomada del Anuario Estadístico del INE y presenta el número de visitantes a los museos de algunas ciudades españolas. Esta misma información se presenta en la tabla 3C.2 pero: (1) se han ordenado las provincias por número de habitantes en lugar de orden alfabético; (2) los datos se han redondeado a miles de habitantes; (3) se han añadido las medias de filas y columnas.

Finalmente en la tabla 3C.3 se ha eliminado el efecto de escala para hacer los datos más homogéneos dividiendo por el número de habitantes en cada provincia. Es claro que la capacidad de transmitir información es mucho mayor en la tabla 3C.3 que en la 3C.1. En resumen, al presentar información mediante una tabla conviene siempre:

1) Escoger cuidadosamente el orden de las filas y las columnas. Cuando exista una secuencia temporal, siempre debe mantenerse. En ausencia de otro criterio, ordenar las filas (columnas) por su tamano medio.
2) Redondear y escoger las unidades de manera que cada dato contenga como máximo tres dígitos.
3) Escribir las medias de filas y columnas.
4) Si existe una variable de escala (tamano) que es importante para explicar la variabilidad de la tabla, dividir cada dato por su variable de escala para hacer homogéneas las comparaciones.

Tabla 3C.1 Visitantes de museos por trimestres en 1987 en algunas ciudades españolas. Fuente, INE

	1T	2T	3T	4T
Albacete	10.601	31.535	5.419	8.180
Badajoz	28.184	82.756	66.833	33.080
Cáceres	14.793	30.686	31.025	15.073
Madrid	686.570	971.912	796.848	666.770
Sevilla	32.672	42.676	24.231	16.170
Toledo	104.441	266.683	170.257	90.700
Valencia	25.222	74.056	50.690	38.601
Valladolid	31.288	77.323	51.951	29.586

Tabla 3C.2 Reordenación y redondeo de los datos de la tabla 3C.1

	1T	2T	3T	4T	Medias
Madrid	687	972	797	667	781
Valencia	25	74	51	39	47
Sevilla	33	43	24	16	29
Badajoz	28	83	67	33	53
Valladolid	31	77	52	30	48
Toledo	104	267	170	91	158
Cáceres	15	31	31	15	23
Albacete	11	32	5	8	14
Medias	117	197	150	112	

Tabla 3C.3 Visitantes de museos por mil habitantes en 1987 en algunas provincias españolas

	1T	2T	3T	4T	Medias
Toledo	220	567	360	193	335
Madrid	145	206	169	141	165
Valladolid	63	157	106	61	97
Badajoz	41	121	98	48	77
Cáceres	36	75	75	36	56
Albacete	33	95	15	24	42
Valencia	12	36	25	19	23
Sevilla	22	29	16	11	20
Medias	72	161	108	67	

Apéndice 3D: Propiedades de la matriz de covarianzas

Carácter no negativo

Vamos a demostrar que la matriz de covarianzas es siempre no negativa (semidefinida positiva), es decir, tanto el determinante como los menores principales son positivos y dado cualquier vector **c** de números reales:

$$\mathbf{c}'\mathbf{M}\mathbf{c} \geq 0 \tag{A.1}$$

donde **M** es la matriz de covarianzas definida por (3.14).

Para demostrar (A.1) definamos una nueva variable v por:

$$v = \mathbf{c}'(\mathbf{X} - \bar{\mathbf{X}}) \tag{A.2}$$

que, por construcción, tendrá media cero. Su varianza es:

$$Var(v) = 1/n \sum v_i^2 = 1/n \sum \mathbf{c}'(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'\mathbf{c}$$

y como la varianza es siempre no negativa

$$\mathbf{c}'[1/n \sum (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})']\mathbf{c}' \geq 0$$

es decir, sea cual sea el vector **c**:

$$\mathbf{c}'\mathbf{M}\mathbf{c} \geq 0$$

que es la condición necesaria y suficiente para que **M** sea semidefinida positiva.

Reducción de la dimensión

Una conclusión importante de esta demostración es que si existe un vector **c** tal que:

$$\mathbf{c}'\mathbf{M}\mathbf{c} = 0 \qquad \text{(A.3)}$$

entonces hay una variable que es una combinación lineal exacta de las demás, lo que equivale a decir que en lugar de k variables tenemos $k - 1$ variables distintas. En efecto, esta condición implica que la varianza de la combinación lineal es cero, y, como su media es cero, la variable debe ser idénticamente nula para todos los puntos, es decir:

$$w_1(x_{1i} - \bar{x}_1) + ... + w_k(x_{ki} - \bar{x}_k) = 0 \qquad i = 1, ..., n \qquad \text{(A.4)}$$

Entonces una variable podrá despejarse de (A.4) a partir de las $k - 1$ variables restantes y quedará determinada por las demás. Por ejemplo, con tres variables si la relación es

$$2(x_1 - 5) + 3x_2 + (x_3 - 2) = 0$$

podemos elegir dos cualesquiera de ellas y obtener la tercera por diferencia. Esto es debido a que todos los coeficientes w_i son no nulos. Por el contrario, en

$$x_2 + 4(x_3 - 2) = 0$$

tendremos que conservar x_1 y elegir entre x_2 y x_3.

La condición (A.3) implica que el rango de la matriz **M** es $k - 1$, en lugar de k, es decir, existe en la matriz **M** una fila que es combinación lineal de las demás. Por tanto, una forma rápida de comprobar si no hay variables redundantes es estudiar el rango de la matriz **M**: si éste es k, las variables son distintas, si es $k - 1$, es posible eliminar una variable. Para identificarla, necesitamos encontrar un vector **c** tal que:

$$\mathbf{M}\mathbf{c} = 0 \qquad \text{(A.5)}$$

es decir, **c** debe ser el vector propio de la matriz **M** asociado al valor propio cero. Los coeficientes no nulos de **c** indicarán qué variables están relacionadas entre sí y, entre ellas, podremos arbitrariamente eliminar una.

Generalizando esta idea, si el rango de la matriz **M** es $k - h$, existen h vectores que verifican (A.3), es decir, h combinaciones lineales nulas que permitirán despejar h variables en función de las demás. El número de variables no redundantes será pues $k - h$, el rango de la matriz. Además, las h ecuaciones se encuentran dadas por los h vectores propios asociados al valor propio cero de la matriz **M**.

En general hay muchas formas distintas de seleccionar las $k - h$ variables no redundantes: podemos escoger un subconjunto de las variables originales, pero también ciertas combinaciones lineales de ellas que tengan buenas propiedades. La representación más simple de un conjunto de variables es cuando éstas tiene covarianzas nulas, ya que entonces su matriz de varianzas y covarianzas es diagonal. En general las variables originales serán dependientes, pero si definimos $k - h$ nuevas variables mediante:

$$v_i = \mathbf{c}_i'(\bar{\mathbf{X}}_i - \bar{\mathbf{X}}) \qquad i = 1, ..., k - h \tag{A.6}$$

donde $\mathbf{c}_i$ es un vector propio asociado a un valor propio no nulo de la matriz **M**, es decir, que verifica la relación:

$$\mathbf{M}\mathbf{c}_i = \lambda_i \mathbf{c}_i \qquad i = 1, ..., h$$

entonces, las $k - h$ nuevas variables (A.6) construidas a partir de las k variables originales tienen media cero, y:

$$Var(v_i) = \mathbf{c}_i' \mathbf{M} \mathbf{c}_i = \lambda_i \mathbf{c}_i' \mathbf{c}_i = \lambda_i$$
$$Cov(v_i v_j) = \mathbf{c}_i' \mathbf{M} \mathbf{c}_j = \lambda_j \mathbf{c}_i' \mathbf{c}_j = 0$$

es decir, tendrán varianzas iguales a los valores propios de **M**, y como los vectores propios de una matriz simétrica son ortogonales, tendrán covarianzas nulas. Estas nuevas variables tienen por tanto la misma varianza generalizada y contienen la misma información que las k originales, pero representan un conjunto más simple.

Este procedimiento puede aplicarse también cuando el rango de **M** es aproximadamente $k - h$, es decir, cuando la matriz tiene h valores propios muy pequeños con relación al resto. Entonces este método permite recoger la información de las k variables mediante un conjunto más simple de $k - h$ variables incorreladas, que son combinación lineal de las originales. Estas nuevas variables se denominan *componentes principales* del conjunto de datos.

Segunda parte

Modelos

4. Probabilidad y variables aleatorias

Andrei Nikolaevich Kolmogorov
(1903-1987)
Matemático ruso, fundador del cálculo de probabilidades moderno. Desde muy joven mostró una extraordinaria aptitud para las matemáticas. Un pionero en muchas ramas de la matemática, estableció el cálculo de probabilidades sobre unos fundamentos axiomáticos precisos.

4.1 Introducción

Cuando los datos que estudiamos son una muestra de una población, el problema central es inferir las propiedades de ésta a partir de la muestra. El instrumento conceptual que permitirá esta generalización es un modelo de la población, es decir, una representación simbólica de su comportamiento. Los modelos estadísticos van a actuar de puente entre lo observado (muestra) y lo desconocido (población). Su construcción y estudio es el objetivo del cálculo de probabilidades.

4.2 Probabilidad y sus propiedades

4.2.1 Concepto

El concepto de probabilidad se aplica a los elementos de una población homogénea. Supongamos una población finita con N elementos, k de los cuales tienen la característica A. Llamaremos probabilidad de la característica A en la población a la frecuencia relativa k/N. Escribiremos:

$$P(A) = \frac{k}{N}$$

Supongamos ahora que intentamos extender este concepto a una población homogénea pero cuyo tamaño es ilimitado. Por ejemplo, observamos el sexo de una persona al nacer, la ocurrencia o no de un accidente o el resultado de tirar una moneda. Un hecho comprobable empíricamente es que la frecuencia relativa de aparición de estos sucesos tiende, al aumentar el número de observaciones, hacia un valor constante. Esta propiedad fue inicialmente descubierta en los juegos de azar: al tirar una moneda, la frecuencia relativa del suceso cara tiende, al aumentar el número de tiradas, hacia el valor constante 1/2 si la moneda está bien hecha (véase la figura 4.1). Posteriormente, se observó esta misma propiedad en datos demográficos (por ejemplo, la frecuencia relativa de nacimiento de varones tiende hacia 0,51), así como en multitud de fenómenos económicos, industriales y sociales.

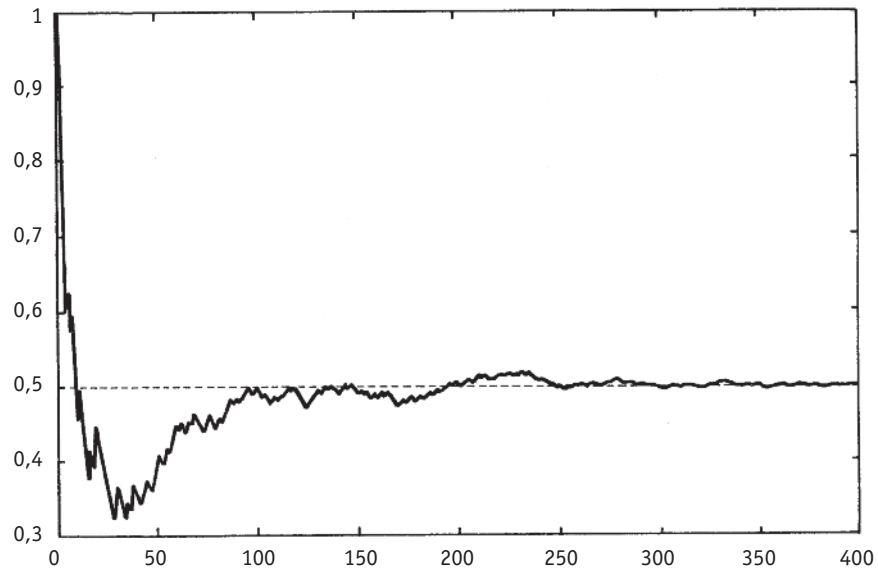
Estas experiencias condujeron en el siglo XIX a definir la probabilidad de un suceso como el valor límite de su frecuencia relativa al repetir indefinidamente la experimentación.

Esta definición presenta problemas importantes: desde el punto de vista teórico el límite anterior no puede interpretarse en el sentido del análisis, ya que no es posible fijar a priori un número de repeticiones n tal que, a partir de él, la diferencia entre la frecuencia relativa y la probabilidad sea menor que una cantidad prefijada; desde el punto de vista práctico la definición implica la imposibilidad en muchos casos de un conocimiento exacto de la probabilidad, ya que:

1) Al no ser posible una experimentación indefinida, la información disponible respecto a la frecuencia relativa es siempre limitada.
2) El sistema observado puede variar a lo largo del tiempo, y con él las frecuencias relativas.

Por tanto, aunque para poblaciones finitas la identificación de probabilidad con la frecuencia relativa es simple y directa, para poblaciones infi-

Figura 4.1 Evolución de la frecuencia relativa de cara al lanzar una moneda 400 veces



nitas presenta problemas importantes. Las dificultades se hacen insalvables al intentar extender el concepto de probabilidad a sucesos inciertos que solamente ocurrirán una vez y donde ni existe ni es posible generar una población de observaciones homogéneas donde calcular la frecuencia relativa. Por ejemplo, la probabilidad de que un nuevo producto tenga éxito, un atleta supere un récord o se produzca un accidente en una central nuclear.

Para evitar estos inconvenientes, la probabilidad se definió en los años treinta axiomáticamente: sus propiedades corresponden a las de la frecuencia relativa, y se encuadran dentro de la teoría general de la medida. La probabilidad sería entonces una medida de incertidumbre, con propiedades similares a las medidas de longitudes, tiempo, etc.

Este enfoque evita la definición conceptual de la probabilidad y, por tanto, no ofrece una guía de cómo calcularla en la práctica. Una concepción más operativa es definir la probabilidad como una medida personal de la incertidumbre de un suceso, basada en aquellas experiencias previas que, con la información disponible, se consideren indistinguibles o intercambiables. Esta medida es forzosamente personal, ya que depende del grado de información. En situaciones repetitivas, cuando exista una amplia experiencia, la probabilidad vendrá determinada por la frecuencia relativa, mientras que, en otros casos, dependerá de distintos tipos de información.

Estrictamente, pues, la probabilidad depende del grado de información disponible, y la probabilidad de un suceso *A* debería indicarse como *P*(*A/I*), donde *I* representa un conjunto de información definida que contiene:

a) Los sucesos posibles al realizar el experimento. Se denomina *espacio muestral* a este conjunto de todos los sucesos posibles que es definido por el experimentador.
b) La evidencia empírica existente respecto a la ocurrencia de estos sucesos.

Para simplificar, supondremos en adelante que el conjunto *I* está perfectamente definido, y escribiremos *P*(*A*) para indicar la probabilidad de un suceso cualquiera.

4.2.2 Definición y propiedades

Población

Una población es un conjunto de elementos homogéneos en los que se desea investigar la ocurrencia de una característica o propiedad. El número de elementos puede ser finito (por ejemplo, los estudiantes matriculados en una universidad) o teóricamente infinito (las personas que hoy y en el futuro subscriban un seguro de vida, las piezas fabricadas por una máquina, la demanda diaria en un supermercado), pero debe ser posible, aunque sea conceptualmente, observar sus elementos. La población debe definirse sin ambigüedad, de manera que sea posible siempre discriminar si un elemento pertenece o no a ella.

El proceso de observar en un elemento de la población la característica o propiedad de interés para el investigador se denomina *experimento.*

Sucesos elementales y compuestos

Llamaremos *sucesos elementales* de un experimento a un conjunto de resultados posibles (*a, b, c, ...*) que verifican:

1) Siempre ocurre alguno de ellos.
2) Son mutuamente excluyentes: la ocurrencia de uno implica la no ocurrencia de los demás.

Llamaremos *sucesos compuestos* a los construidos a partir de uniones de resultados elementales. Por ejemplo:

Experimento	Sucesos elementales	Sucesos compuestos
Tirar un dado	(1, 2, 3, 4, 5, 6)	Número par; número impar; menor que 4; múltiplo de 3.
Contar los varones en familias con tres hijos	(0, 1, 2, 3)	Más de uno; menos de tres.
Contar el número de averías de una máquina en un mes	(0, 1, ..., 30)	Más de 10; menos de 20; entre 5 y 15 inclusive.

En el primer ejemplo los elementos de la población son las sucesivas tiradas de un dado; en el segundo, las familias (españolas, urbanas, etc.) que tienen exactamente tres hijos; en el tercero, los meses de trabajo de una máquina en condiciones definidas.

Llamaremos *espacio muestral* al conjunto de resultados posibles del experimento. Los sucesos (elementales o compuestos) son, por tanto, subconjuntos del espacio muestral. Por conveniencia consideraremos como suceso al mismo espacio muestral y lo llamaremos suceso seguro, E, porque siempre ocurre. También incluiremos en el conjunto de sucesos el suceso imposible, $\varnothing$, que no ocurre nunca. Se desea asociar a cada suceso una medida de incertidumbre que llamaremos probabilidad con las propiedades siguientes.

Propiedades:

1) La frecuencia relativa de un suceso A, $fr(A)$, es un valor entre cero y uno, por tanto:

$$0 \leqslant P(A) \leqslant 1 \tag{4.1}$$

2) La frecuencia relativa del suceso seguro, E, que ocurre siempre, es uno, y por tanto:

$$P(E) = 1 \tag{4.2}$$

3) Si A y B son categorías mutuamente excluyentes y las unimos en una nueva $C = A + B$, que ocurre cuando se da o bien A o bien B, la frecuencia relativa de C es la suma de las frecuencias relativas de A y B. Por tanto, para sucesos mutuamente excluyentes:

$$P(A + B) = P(A) + P(B) \tag{4.3}$$

4) Si A y B no son mutuamente excluyentes y llamamos n_{AB}, $n_{A\bar{B}}$, $n_{\bar{A}B}$ al número de veces que aparecen los sucesos mutuamente excluyentes: (A y B), (A y no B), (no A y B), tendremos:

$$n_A = n_{AB} + n_{A\bar{B}}$$
$$n_B = n_{AB} + n_{\bar{A}B}$$
$$n_{A+B} = n_{AB} + n_{A\bar{B}} + n_{\bar{A}B}$$

de donde se obtiene la siguiente relación:

$$n_{A+B} = n_A + n_B - n_{AB}$$

y dividiendo por el número total de observaciones resulta una relación entre frecuencias relativas que traducida a probabilidades es:

$$P(A + B) = P(A) + P(B) - P(AB) \tag{4.4}$$

5) Si $\bar{A}$ es el suceso complementario de A, que ocurre siempre que no lo hace A, las propiedades (4.2) y (4.3) implican que:

$$P(\bar{A}) = 1 - P(A) \tag{4.5}$$

Una conclusión de esta propiedad es que la probabilidad del suceso complementario del suceso seguro —que llamaremos suceso imposible— será cero.

4.2.3 La estimación de probabilidades en la práctica

En la práctica no tiene sentido hablar de probabilidades sin definir previamente la población a la que nos referimos y los sucesos que vamos a considerar.

La determinación de probabilidades para sucesos compuestos requiere conocer las de los sucesos elementales. Estas probabilidades se determinan:

1) Estudiando la frecuencia relativa al repetir el experimento en condiciones similares. Este método sólo es factible en ocasiones en que es posible una experimentación continuada.
2) Encontrando, a partir de la naturaleza del experimento, relaciones que liguen sus probabilidades elementales y determinen sus valores. El caso más simple es el de equiprobabilidad, que estudiaremos a continuación.
3) Combinando la experimentación con la teoría sobre la naturaleza del experimento. Éste es el método más frecuente en la práctica y más fructífero. Lo utilizaremos en el capítulo 5 para construir los modelos de distribución de probabilidad más importantes.

El caso de equiprobabilidad

En ocasiones, la simetría de los sucesos elementales sugiere considerarlos equiprobables. Este razonamiento se ha aplicado repetidamente en los juegos de azar a problemas como tirar dados o monedas, extraer naipes de barajas, etc. A veces, el mecanismo generador de los resultados está diseñado para intentar asegurar esta equiprobabilidad, como en la lotería o la ruleta.

En estos casos, si existen n sucesos elementales equiprobables, la probabilidad de cada uno de ellos debe ser $1/n$, para asegurar que la suma total sea uno. La probabilidad de un suceso compuesto A que contiene f sucesos elementales será f/n, lo que da lugar a la regla:

$$P(A) = \frac{\text{casos favorables } (f)}{\text{casos posibles } (n)}$$

Esta regla sólo debe utilizarse cuando la simetría esté confirmada por el mecanismo generador (como en la lotería) o por la evidencia empírica.

Ejemplo 4.1

En una estación existen cuatro taquillas servidas por cuatro personas igualmente eficientes. Un estudiante llega cada día y se coloca en una de las cuatro colas. Calcular la probabilidad de que la cola escogida no sea la más rápida.

Solución

Llamando 1, 2, 3, 4 a las cuatro colas, sea a_i el suceso la cola i (i = 1, 2, 3, 4) es la más rápida. Si despreciamos la posibilidad de empate, el espacio muestral está formado por los sucesos elementales (a_1, a_2, a_3 y a_4). Por hipótesis, en promedio las colas son igualmente rápidas, por lo que:

$$P(a_1) = \ldots = P(a_4) = 1/4$$

Supongamos que el estudiante escoge la cola i. Entonces la probabilidad de que esta cola sea la más rápida es 1/4. El suceso complementario $\bar{a}_i$, que haya una cola más rápida, será por tanto:

$$P(\bar{a}_1) = 1 - P(a_i) = 3/4$$

Por tanto, el 75% de los días el estudiante observará una cola más rápida que la suya (si es pesimista, encontrará razones para pensar que tiene la mala suerte de escoger siempre la más lenta).

4.3 Probabilidad condicionada

4.3.1 Concepto

La frecuencia relativa de A condicionada a la ocurrencia de B se define considerando únicamente los casos en los que aparece B, y viendo en cuántos de estos casos ocurre el suceso A; es, por tanto, igual a la frecuencia de ocurrencia conjunta de A y B, partida por el número de veces que ha ocurrido B. Escribiremos:

$$fr(A|B) = \frac{n_{AB}}{n_B}$$

entonces, como $fr(A) = n_A/n$; $fr(B) = n_B/n$; $fr(AB) = n_{AB}/n$, se tiene:

$$fr(A|B) = \frac{fr(AB)}{fr(B)}$$

o, lo que es lo mismo,

$$fr(AB) = fr(A|B)fr(B) = fr(B|A)fr(A)$$

En consecuencia, exigiremos esta misma propiedad a la probabilidad y definiremos *probabilidad de un suceso A condicionada* a otro *B* por:

$$P(A|B) = \frac{P(AB)}{P(B)} \tag{4.6}$$

donde AB representa el suceso ocurrencia conjunta de A y B, y suponemos $P(B) > 0$.

Es importante diferenciar entre $P(AB)$ *y* $P(A|B)$. El primer término indica la probabilidad de que ocurran conjuntamente los sucesos A y B y *siempre* es menor que la probabilidad de A o de B. El segundo es la probabilidad de que en los casos en que ya ha ocurrido B ocurra también A, y puede ser mayor, menor o igual que $P(A)$. En el primer caso nos movemos dentro del espacio muestral original, mientras que en el segundo el espacio muestral es el suceso B. Por ejemplo, si lanzamos una moneda dos veces y A es el suceso cara en la primera tirada y B el suceso cara en la segunda, $P(AB)$ está definida en el espacio muestral: $\{AB, A\overline{B}, \overline{A}B, \overline{A}\overline{B}\}$ y tiene probabilidad 1/4; $P(B|A)$ lo está en el espacio $\{B, \overline{B}\}$ y tiene probabilidad 1/2.

Ejemplo 4.2

En una universidad la proporción de graduados en Ciencias Sociales es del 50%, en Ciencias Naturales el 30% y en Ingenierías el 20%. La proporción de mujeres graduadas es del 45% y su distribución entre las titulaciones es 60% Ciencias Sociales, 30% Naturales y 10% Ingenierías. ¿Cuál es la probabilidad de que un graduado en Ciencias Sociales sea mujer? ¿Cuál es la probabilidad de que un hombre sea graduado en Ciencias Naturales?

Sean $\{M,H\}$ los sucesos mujer y hombre y $\{S,N,I\}$ los sucesos graduación en Ciencias Sociales, Naturales o Ingeniería. El espacio muestral para el problema es $\{MS, HS, MN, HN, MI, HI\}$. Si representamos la probabilidad del suceso seguro por un cuadrado de lado unidad (y por tanto superficie igual a uno), las probabilidades de los sucesos pueden asociarse a superficies dentro de este cuadrado. Tomando rectángulos para simplificar, la probabilidad p de un suceso se representará mediante un rectángulo de dimensiones a y b tales que $a, b = p$. Por ejemplo, las probabilidades de los sucesos S,N,I se representarán como rectángulos de base 0,5, 0,3 y 0,2 y altura unidad.

La probabilidad de que una persona sea mujer y graduada en Ciencias Sociales, $P(MS)$, se calcula por $P(S|M)P(M)$. Según los datos del problema $P(S|M) = 0{,}6$ y $P(M) = 0{,}45$, por lo que $P(MS) = 0{,}6 \cdot 0{,}45 = 0{,}27$. También, $P(MN) = P(N|M)P(M) = 0{,}3 \cdot 0{,}45 = 0{,}135$ y $P(MI) = P(I|M) \cdot P(M) = 0{,}10 \cdot$

· 0,45 = 0,045. Para representar gráficamente estas probabilidades utilizaremos que al ser $P(S) = P(MS) + P(HS)$, la probabilidad $P(S)$ debe dividirse en dos partes: $P(MS)$, de área 0,27, y $P(HS)$, de área:

$$P(HS) = P(S) - P(MS) = 0,5 - 0,27 = 0,23$$

Como $P(S)$ es un rectángulo de base 0,5 y altura unidad, la altura del rectángulo de base 0,5 asociado a $P(MS)$ deberá elegirse para que el área sea 0,27, con lo que, llamando h a esta altura, su valor se calculará así:

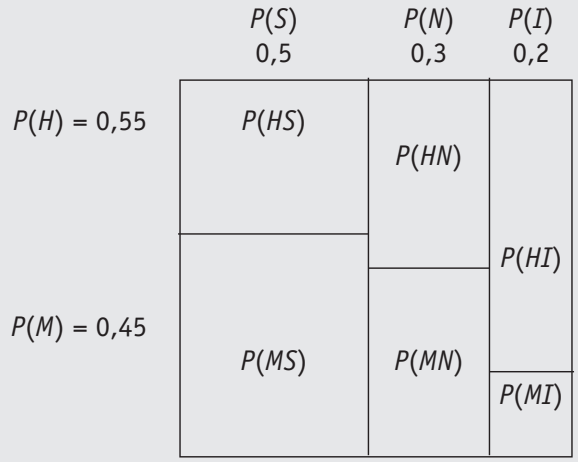
$$0,27 = 0,5 \times h$$

Por tanto $h = 0,27/0,5 = 0,54$ y $P(MS)$ será un rectángulo de base 0,5 y altura 0,54. Análogamente, $P(HS)$ será el rectángulo de base 0,5 y altura 0,46 (0,23/0,5).

Para interpretar esta operación observemos que hemos asignado el 54% del área de $P(S)$ a $P(MS)$ y el 46% a $P(HS)$. La probabilidad condicionada de que un graduado en Ciencias Sociales sea mujer será:

$$P(M|S) = \frac{P(MS)}{P(S)} = \frac{0,27}{0,5} = 0,54$$

Por tanto la altura h del rectángulo que hemos calculado para $P(MS)$ para incluirlo dentro de $P(S)$ es la probabilidad condicionada de $P(M|S)$ de que un graduado en Ciencias Sociales sea mujer.



Del mismo modo, las probabilidades de los sucesos MN, HN se representan dividiendo la probabilidad $P(N)$ en dos partes. Como $P(MN) = 0,135$ y

$P(MI) = 0{,}045$, $P(M|N) = 0{,}135/0{,}3 = 0{,}45$ y $P(M|I) = 0{,}045/0{,}2 = 0{,}225$, que serán las alturas de los rectángulos asociadas a las probabilidades $P(MN)$ y $P(MI)$. Observemos que la relación $P(MN) = P(M|N)\,P(N)$ equivale a «área = altura · base».

La probabilidad de que un hombre graduado lo sea en Ciencias Naturales se obtendrá calculando la proporción que representa $P(HN)$ sobre el total de $P(H)$, es decir, $P(N|H) = P(HN)/P(H)$. Como $P(HN) = P(N) - P(MN) = 0{,}3 - 0{,}135 = 0{,}165$, y $P(H) = 1 - P(M) = P(HS) + P(HN) + P(HI) = 0{,}55$, tendremos

$$P(N|H) = \frac{0{,}165}{0{,}55} = 0{,}3$$

Cualquier otra probabilidad se calcula análogamente. Cuando el espacio muestral puede dividirse en sucesos mediante dos criterios de clasificación, como en este caso: sexo y titulación, la representación gráfica que hemos utilizado ayuda a interpretar el significado intuitivo de las probabilidades condicionadas y conjuntas. Esta idea puede extenderse a situaciones más generales.

El diagrama muestra que las alturas $P(M|S)$, $P(M|N)$ y $P(M|I)$ son distintas entre sí y distintas de $P(M)$. Intuitivamente esto sugiere que el suceso mujer M y el suceso S, N o I están relacionados. Esto conduce al concepto de independencia que presentamos a continuación.

4.3.2 Independencia de sucesos

Diremos que dos sucesos A y B son *independientes* si el conocimiento de la ocurrencia de uno no modifica la probabilidad de aparición del otro. Por tanto, A y B son independientes si:

$$P(A|B) = P(A)$$

$$P(B|A) = P(B)$$

Por (4.6), una definición equivalente de independencia de dos sucesos es:

$$\boxed{P(AB) = P(A)P(B)} \qquad (4.7)$$

Esta definición se generaliza para cualquier número de sucesos: diremos que los sucesos $A_1, \ldots, A_n$ son independientes si la probabilidad conjunta de

cualquier subconjunto que pueda formarse con ellos es el producto de las probabilidades individuales.

La independencia entre sucesos puede en algunos casos preverse, pero en general debe determinarse experimentalmente. Por ejemplo, las averías en dos talleres contiguos pueden ser independientes si éstos no guardan relación, y dependientes si las averías van ligadas al tipo de producto fabricado y ambos talleres producen el mismo.

Ejemplo 4.3

Calcularemos la probabilidad de que una familia con tres hijos tenga más de un varón. El espacio muestral, llamando *V* a varón y *M* a mujer, contiene los ocho sucesos elementales {*VVV*, *VVM*, *VMV*, *MVV*, *VMM*, *MVM*, *MMV*, *MMM*}. El suceso *A*, más de un varón, es la unión de los cuatro resultados elementales: *A* = {*VVV*, *VVM*, *VMV*, *MVV*}, y la probabilidad pedida será la suma de las probabilidades de estos cuatro sucesos elementales. Comenzando con el primero, $P(VVV) = P(V)P(V/V)P(V/VV)$, y si suponemos *independencia* entre nacimientos —que habría que comprobar experimentalmente viendo si la frecuencia relativa de nacimiento de varón después de varón es igual a la frecuencia relativa de varón—, entonces: $P(VVV) = P(V)^3$, y $P(VVM) = P(VMV) = P(MVV) = P(V)^2P(M)$.

Puede argumentarse que, por simetría, $P(V)$ debe ser igual a $P(M)$. La experiencia empírica nos dice que esto *no* es cierto, ya que la frecuencia relativa de varón es 0,51 y la de mujer 0,40. Por tanto: $P(A) = 0{,}51^3 + 3 \cdot \cdot 0{,}51^2 \cdot 0{,}49$.

Ejemplo 4.4

Se denomina *fiabilidad* de un sistema a la probabilidad de que funcione satisfactoriamente. Supongamos un sistema S_1 (eléctrico, mecánico, humano...) formado por 50 componentes que deben funcionar todos correctamente para que lo haga el sistema. La probabilidad de que cada componente funcione después de 100 horas es 0,99, y los componentes se averían independientemente. ¿Cuál es la fiabilidad del sistema después de 100 horas?

Sea A_i el suceso; el componente *i* funciona, y $\bar{A}_i$, no funciona. Entonces:

$$\text{Fiabilidad} = P(A_1A_2, \ldots, A_{50}) = P(A_1) \ldots (A_{50}) = 0{,}99^{50} = 0{,}605$$

Este ejemplo ilustra un resultado general: en sistemas complejos, aunque la fiabilidad de cada componente sea alta, la fiabilidad del sistema puede ser baja. Para aumentar la fiabilidad, podemos disponer varios siste-

mas en paralelo de manera que el sistema conjunto funcione si uno de los sistemas individuales lo hace. Para concretar, supongamos que se trata de un sistema de seguridad que se duplica, es decir, el doble sistema (S^1 + + S_2) funciona si una de las dos cadenas de 50 elementos funciona. Entonces, aplicando (4.4):

$$\text{Fiabilidad} = P(\text{funcione } S_1 \text{ o } S_2) = P(S_1) + P(S_2) - P(S_1S_2)$$

si los dos sistemas son independientes: $P(S_1S_2) = P(S_1)P(S_2) = 0{,}366$, y

$$\text{Fiabilidad} = 2 \cdot (0{,}605) - 0{,}605^2 = 0{,}844$$

Obtendríamos el mismo resultado sumando las probabilidades de los sucesos disjuntos $\bar{S}_1S_2$, $S_1\bar{S}_2$, S_1S_2 que forman el suceso compuesto: el sistema funciona. Entonces,

$$\text{Fiabilidad} = (0{,}395)(0{,}605) + (0{,}605)(0{,}395) + (0{,}605)^2 = 0{,}844$$

Finalmente, podríamos también resolver el problema calculando la probabilidad de que *no* funcione; $P(\text{no funcione}) = P(\bar{S}_1)P(\bar{S}_2) = 0{,}395^2 = 0{,}156$, entonces la fiabilidad es la probabilidad del suceso complementario:

$$\text{Fiabilidad} = 1 - 0{,}156 = 0{,}844$$

4.3.3 Teorema de Bayes

Consideremos un experimento que se realiza en dos etapas: en la primera, los sucesos posibles, $A_1, \ldots, A_n$, son mutuamente excluyentes, con probabilidades conocidas, $P(A_i)$, y tales que:

$$\Sigma P(A_i) = 1$$

En la segunda etapa, los resultados posibles, B_j, dependen de los de la primera, y se conocen las probabilidades condicionadas $P(B_j|A_i)$ de obtener cada posible resultado B_j cuando aparece en la primera etapa el A_i.

Se efectúa ahora el experimento, pero el resultado de la primera fase, A_i, no se conoce, aunque sí el de la segunda, que resulta ser B_j. El teorema de Bayes permite calcular las probabilidades $P(A_i|B_j)$ de los sucesos no observados de la primera etapa, dado el resultado observado en la segunda.

Partiendo de la definición de probabilidad condicionada:

$$P(A_i|B_j) = \frac{P(A_iB_j)}{P(B_j)} = \frac{P(B_j|A_i)P(A_i)}{P(B_j)}$$

y, por otro lado:

$$P(B_j) = P(B_jA_1 + B_jA_2 + ... + B_jA_n) \tag{4.8}$$

ya que B_j debe ocurrir con alguno de los n posibles sucesos A_i. Como los sucesos B_jA_1, B_jA_2, ... son mutuamente excluyentes, al serlo los A_i, tenemos:

$$P(B_j) = \sum_i P(B_jA_i) = \sum_i P(B_j|A_i)P(A_i)$$

y sustituyendo en la expresión de $P(A_i/B_j)$:

$$P(A_i|B_j) = \frac{P(B_j|A_i)P(A_i)}{\sum_i P(B_j|A_i)P(A_i)} \tag{4.9}$$

que se conoce como teorema de Bayes.

Ejemplo 4.5

Se dispone de dos urnas. La urna U_1 contiene el 70% de bolas blancas y el 30% de bolas negras, y la U_2, el 30% de bolas blancas y el 70% de bolas negras. Se selecciona una de estas urnas al azar y se toman diez bolas una tras otra con reemplazamiento. El resultado es: B = *bnbbbbnbbb*, donde *b* indica bola blanca y *n* negra. Se pregunta: ¿Cuál es la probabilidad de que esta muestra provenga de U_1?

Este experimento puede suponerse como incluyendo dos etapas: la primera es seleccionar la urna (U_1, U_2), y la segunda, la muestra dentro de la urna. Como hay dos urnas y se toma una al azar:

$$P(U_1) = P(U_2) = \frac{1}{2}$$

El suceso B está compuesto por la ocurrencia conjunta de 10 sucesos independientes, ya que el resultado de una extracción con reemplazamiento no modifica las probabilidades de las siguientes. Como:

$$P(b|U_1) = 0{,}7 \qquad P(n|U_1) = 0{,}3$$

se verifica:

$$P(B|U_1) = P(bnbbbbnbbb|U_1) = P(b|U_1) \cdot P(n|U_1) \cdot P(b|U_1) \ldots P(b|U_1) = \\ = 0{,}7^8 \cdot 0{,}3^2$$

Análogamente:

$$P(B|U_2) = 0{,}3^8 \cdot 0{,}7^2$$

La probabilidad pedida es $P(U_1|B)$. Aplicando el teorema de Bayes:

$$P(U_1|B) = \frac{P(B|U_1)P(U_1)}{P(B|U_1)P(U_1) + P(B|U_2)P(U_2)}$$

$$P(U_1|B) = \frac{0{,}7^8 \cdot 0{,}3^2 \cdot \left(\frac{1}{2}\right)}{\left(\frac{1}{2}\right) 0{,}7^8 \cdot 0{,}3^2 + \left(\frac{1}{2}\right) 0{,}3^8 \cdot 0{,}7^2} = \frac{0{,}7^6}{0{,}7^6 + 0{,}3^6} = 0{,}994$$

Por tanto, el resultado B proporciona una alta seguridad de que la muestra se ha extraído de la urna U_1.

Discusión: Al presentar este problema a un grupo de 100 estudiantes y pedirles que estimasen la probabilidad pedida, se obtuvo el siguiente resultado: el 2% estimó un valor entre 0,5 y 0,6; el 20%, entre 0,6 y 0,7; el 60%, entre 0,7 y 0,8; el 15%, entre 0,8 y 0,9; y el 3%, entre 0,9 y 0,95. Ninguna persona supuso un valor mayor que 0,95. Este ejemplo indica una frecuente falta de intuición ante la incertidumbre que tiene consecuencias negativas, ya que desperdiciamos la información obtenida en la experimentación. (Aplíquese a un médico que decide en base a una prueba B entre dos enfermedades U_1, U_2, a un ingeniero que trata de distinguir entre dos causas de averías [U_1, U_2] dados los resultados B, a un científico que selecciona entre dos teorías o hipótesis científicas [U_1, U_2] ante un cuerpo de evidencia empírica.)

Ejemplo 4.6 (Agradezco este ejemplo a A. Maravall.)

Un concursante debe elegir entre tres puertas, detrás de una de las cuales se encuentra un premio. Hecha la elección y antes de abrir la puerta, el presentador le muestra que en una de las dos puertas no escogidas no está el premio y le da la posibilidad de reconsiderar su decisión. ¿Qué debe hacer el concursante?

Definamos los sucesos siguientes:

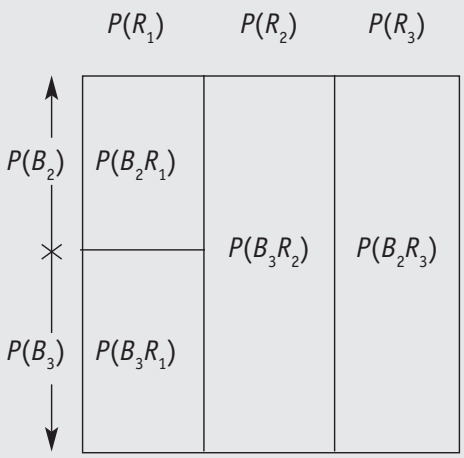
A_i = el concursante elige inicialmente la puerta i; $i = 1, 2, 3$
R_i = el premio realmente está en la i; $i = 1, 2, 3$

El espacio muestral está formado por los nueve sucesos (A_iR_j), cada uno de ellos con probabilidad 1/9. Si, por ejemplo, se da A_1, la probabilidad de ganar es:

$$P(R_1|A_1) = \frac{P(R_1A_1)}{P(A_1)} = \frac{1/9}{1/3} = \frac{3}{9} = \frac{1}{3}$$

Supongamos ahora que un concursante ha escogido la puerta A_1 y haremos todo el análisis condicionado a dicho resultado, aunque por simplicidad no lo indicaremos en la notación. Sea B_i = el presentador abre la puerta i y muestra que no contiene el premio. Según el enunciado, si el concursante ha elegido A_1 el espacio muestran está formado por los cuatro sucesos $\{B_2R_1, B_2R_3, B_3R_1, B_3R_2\}$.

Podemos representar gráficamente las probabilidades de los sucesos $\{R_iB_j\}$ cuando el concursante ha escogido la puerta A_1 como indica el diagrama.



En efecto $P(R_1) = P(R_2) = P(R_3) = 1/3$. Cuando el premio está en la puerta elegida, R_1, tan probable es que el presentador muestre la puerta 2 como la 3, luego $P(B_2|R_1) = P(B_3|R_1) = 1/2$. En consecuencia $P(R_1B_2) = P(R_1B_3) = 0{,}5 \cdot 1/3 = 1/6$. Cuando el concursante elige A_1, y el premio está en la puerta 2, el presentador debe mostrar la puerta 3; luego $P(B_3|R_2) = 1$ y $P(B_3R_2) = P(B_3|R_2)P(R_2) = 1 \cdot 1/3 = 1/3$. Finalmente, cuando el concursante elige A_1 y el premio está en la puerta 3, el presentador mostrará la puerta 2; luego $P(B_2|R_3) = 1$ y $P(B_2R_3) = 1/3$.

La probabilidad de ganar de los concursantes que no cambian de puerta es 1/3. En efecto, si consideramos un concursante que elige la puerta 1, suceso A_1, y el presentador muestra la puerta j (j = 2,3), entonces:

$$P(R_1|B_j) = \frac{P(B_j|R_1)\ P(R_1)}{\Sigma P(B_j|R_i)\ P(R_i)} = \frac{0{,}5 \times \frac{1}{3}}{\frac{1}{3} \times 0{,}5 + 1 \times \frac{1}{3}} = \frac{1}{3}$$

La probabilidad de ganar cambiando de puerta es igual a la probabilidad de que el premio esté en la puerta que *no* muestra el presentador. Suponiendo que muestra la 3, se obtiene:

$$P(R_2|B_3) = \frac{P(B_3|R_2)\ P(R_2)}{\Sigma P(B_3|R_i)\ P(R_i)} = \frac{1 \times \frac{1}{3}}{0{,}5 \times \frac{1}{3} + 1 \times \frac{1}{3}} = \frac{2}{3}$$

Análogamente se comprueba que si muestra la puerta 2, $P(R_3|B_2) = 2/3$.

La razón de que sea conveniente cambiar es que el suceso B_j *no* es independiente de los sucesos R_i, *es decir, el suceso* B_j *da información sobre los* R_i. *En efecto,* $P(B_2) = P(B_3) = 1/2$ y $P(R_1) = P(R_2) = P(R_3) = 1/6$, pero en general $P(B_jR_j) \neq 1{,}6$. Cuando se da A_1, los sucesos R_1 y B_j (j = 2,3) *sí* son independientes, ya que $P(R_1B_2) = P(R_1B_3) = 1/6$, pero los sucesos R_i (i = = 2,3) B_j (j = 2,3) son dependientes como hemos mostrado. Esta dependencia (información) conduce a que convenga reconsiderar la decisión y cambiar de puerta siempre. Si el lector se sorprende por este resultado, considere el siguiente razonamiento: supongamos que este juego se repite muchas veces y que los concursantes siempre cambian su decisión después de mostrarles la puerta vacía. Entonces 2/3 de las veces el premio *no* estaba en su primera elección y al cambiar ganan. Únicamente 1/3 de las veces el premio estaba en su primera elección y pierden al cambiar. En resumen, la probabilidad de ganar si no cambian es 1/3, mientras que si cambian es el doble.

Ejercicios 4.1

4.1.1. Una urna contiene cinco bolas numeradas 1, 2, 3, 4, 5; se pide la probabilidad de que al sacar dos bolas sin reposición la suma de los puntos sea impar.

4.1.2. Las máquinas M_1, M_2 *y* M_3 fabrican en serie piezas similares. Las producciones son de 300, 450 y 600 piezas por hora, y los porcentajes de defectuosas del 2%, 3,5% y 2,5% respectivamente. De la producción total de las tres máquinas reunidas en un almacén al fin de la jornada se toma una pieza al azar. Calcular la probabilidad de que sea defectuosa.

4.1.3. Cuatro fichas están marcadas con las letras *A*, *B*, *C*, *ABC*; se toma una de ellas al azar. Se pregunta si los tres sucesos consistentes en la presencia de la letra *A*, la letra *B* o la *C* sobre la ficha son o no independientes.

4.1.4. En una clase el 30% de los alumnos varones y el 10% de las mujeres son repetidores. El 60% de los alumnos son varones. Si se selecciona un estudiante al azar y resulta repetidor, calcular la probabilidad de que sea mujer.

4.1.5. Tres máquinas M_1, M_2, M_3 fabrican en serie piezas, siendo sus producciones horarias 2.000, 1.000 y 1.000 piezas, y sus fracciones defectuosas 0,05, 0,10 y 0,15. De la producción de un día se toman dos piezas al azar y resultan ambas buenas. Calcular la probabilidad de que ambas procedan de la misma máquina.

4.1.6. Lance un dado 100 veces y estudie la evolución de la frecuencia relativa de cada cara con el número de tiradas.

4.1.7. ¿Pueden ser independientes dos sucesos mutuamente excluyentes que tienen probabilidad no nula?

4.1.8. Tres personas comparten una oficina con un teléfono. De las llamadas que llegan, 2/5 son para *A*, 2/5 para *B* y 1/5 para *C*. El trabajo de estos hombres les obliga a frecuentes salidas, de manera que *A* está fuera el 50% de su tiempo, y *B* y *C* el 25%. Calcular la probabilidad de que:
a) No esté ninguno para responder al teléfono.
b) Esté la persona a la que se llama.
c) Haya tres llamadas seguidas para una persona.
d) Haya tres llamadas seguidas para tres personas diferentes.
Indique las hipótesis realizadas para resolver este problema.

4.1.9. En una clase hay *N* personas. Calcular la probabilidad de que al menos dos tengan el mismo cumpleaños. Indicar las hipótesis realizadas para resolver el problema.

4.1.10. La probabilidad de que un componente de una máquina se averíe antes de 100 horas es 0,01. La máquina tiene 50 componentes; calcular la probabilidad de avería de la máquina antes de 100 horas en los casos siguientes:
1) La máquina se avería cuando lo hace uno o más componentes.
2) La máquina se avería cuando fallan dos o más componentes.
3) La máquina sólo se avería cuando lo hacen todos los componentes.

4.1.11. Un proceso de fabricación puede estar ajustado o desajustado. Cuando está ajustado, produce un 1% de piezas defectuosas, y cuando está desajustado, un 10%. La probabilidad de desajuste es 0,3. Se toma una muestra de diez piezas y todas son buenas. Calcular la probabilidad de que el proceso esté desajustado.

4.1.12. Calcular cuál es el número mínimo de personas a las que usted debe preguntar para que la probabilidad de encontrar una con su mismo cumpleaños sea, al menos, 0,5.

4.1.13. En un campeonato de tenis usted tiene la opción de escoger la secuencia de partidos *A-B-A* o la *B-A-B*, donde *A* y *B* indican sus oponentes. Para clasificarse debe usted ganar dos partidos consecutivos. El jugador *A* es mejor que el *B*. ¿Qué secuencia será preferida?

4.1.14. Un jurado de tres miembros que decide por mayoría tiene dos miembros que deciden independientemente el veredicto correcto con probabilidad p y el tercero lanza una moneda. Si un juez tiene probabilidad p, ¿cuál de los dos sistemas tiene mayor probabilidad de acertar?

4.1.15. Repetir 4.1.14 suponiendo que los tres miembros del jurado tienen ahora probabilidad p. ¿Cuál debe ser el valor de p para que el jurado sea superior al juez individual?

4.1.16. Dos personas tienen que elegir separadamente un número, sabiendo que si eligen el mismo obtendrán un premio. Si usted fuese uno de ellos, ¿que número elegiría?

4.1.17. Calcular la probabilidad de que al extraer cinco cartas de una baraja de póquer (con 52 cartas) se obtenga:
a) Al menos una pareja.
b) Dos parejas (doble pareja).
c) Al menos tres cartas iguales (trío).
d) Una pareja y un trío (full).
e) Cuatro cartas iguales (póquer).

4.4 Variables aleatorias

El cálculo de probabilidades utiliza variables numéricas que se denominan aleatorias, porque sus valores vienen determinados por el azar. En todo proceso de observación o experimento podemos definir una variable aleatoria asignando a cada resultado del experimento un número:

a) Si el resultado del experimento es numérico porque contamos o medimos, los posibles valores de la variable coinciden con los resultados del experimento.
b) Si el resultado del experimento es cualitativo, hacemos corresponder a cada resultado un número arbitrariamente; por ejemplo, 0, si un elemento es bueno, y 1, si es defectuoso.

Diremos que se ha definido una variable aleatoria o que se ha construido un modelo de distribución de probabilidad cuando se especifican los posibles valores de la variable con sus probabilidades respectivas.

4.4.1 Variables aleatorias discretas

Diremos que una *variable aleatoria es discreta* cuando toma un número de valores finito, o infinito numerable. Estas variables corresponden a experimentos en los que se cuenta el número de veces que ha ocurrido un suceso. La distribución de la variable suele definirse mediante la función de probabilidad o la de distribución.

Función de probabilidad

El procedimiento más común de definir una variable aleatoria discreta es indicando sus valores posibles (espacio muestral) y sus probabilidades respectivas. Llamaremos función de probabilidad, $p(x)$, a la función que indica las probabilidades de cada posible valor. Escribiremos;

$$\boxed{p(x_i) = P(x = x_i)} \tag{4.10}$$

Llamando S al espacio muestral, se verificará:

$$\sum_{i \in S} p(x_i) = 1$$

El ejemplo más simple de una variable aleatoria discreta es la *uniforme*, entre 1 y N, cuyo espacio muestral es el conjunto (1, 2, ..., N) y la probabilidad de todos los sucesos es la misma, $p(x_i) = 1/N$.

Función de distribución

Una forma equivalente de caracterizar la distribución de una variable es mediante la *función de distribución*, $F(x)$, definida en cada punto x_0 como la probabilidad de que la variable aleatoria x tome un valor menor o igual que x_0. Escribiremos:

$$F(x_0) = P(x \leq x_0) \tag{4.11}$$

La función de distribución, que se define para todo punto del eje real, es siempre no decreciente, y por convenio:

$$F(-\infty) = 0$$
$$F(+\infty) = 1$$

Suponiendo que la variable x toma los valores posibles ($x_1 \leq x_2 \leq x_3 \ldots \leq \leq x_n$), la función de distribución vendrá definida por:

$$F(x_1) = P(x \leq x_1) = p(x_1)$$
$$F(x_2) = P(x \leq x_2) = p(x_1) + p(x_2)$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$F(x_n) = P(x \leq x_n) = \sum_{i=1}^{n} p(x_i) = 1$$

Por tanto, la función de distribución, $F(x)$, tiene saltos en los puntos de probabilidad no nula del espacio muestral, de magnitud igual a la probabilidad de dicho punto, y es constante en los intervalos entre los puntos de salto.

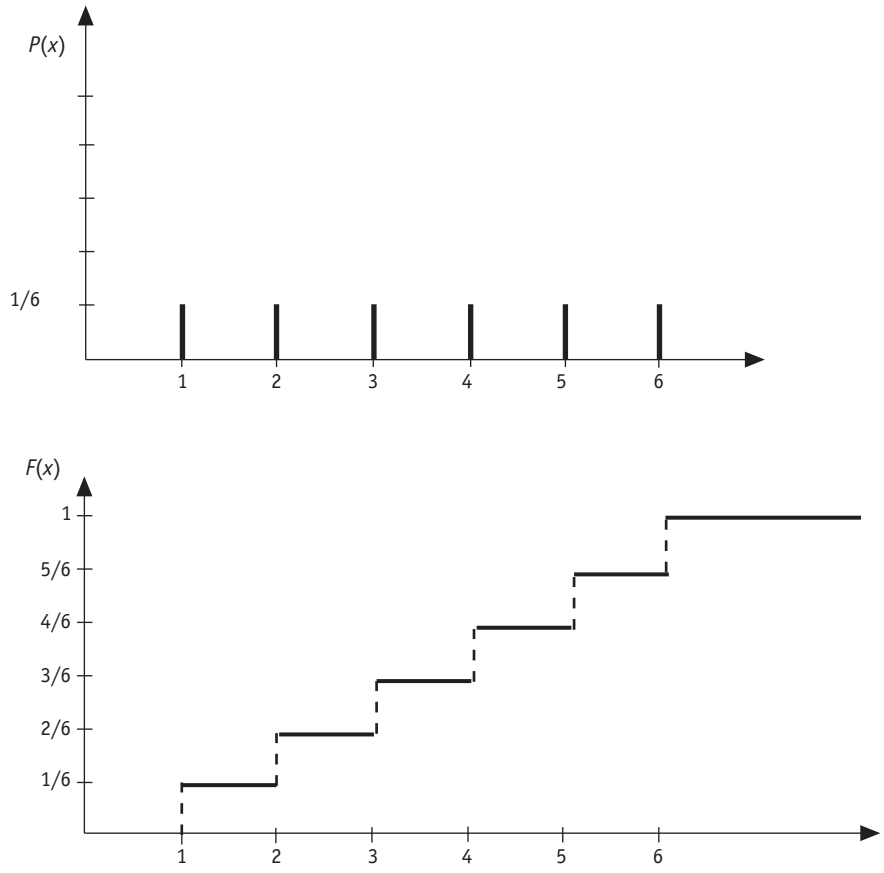
La figura 4.2 representa gráficamente la función de probabilidad y la de distribución para una variable discreta uniforme.

Ejemplo 4.7

Se tira un dado y se define la variable aleatoria: puntuación obtenida. Representar esta variable.

El espacio muestral es {1, 2, 3, 4, 5, 6} y $p(x_i) = 1/6$. Se dice que esta variable es uniforme sobre su espacio muestral. La figura 4.2 presenta su función de probabilidad y de distribución.

Figura 4.2 Función de probabilidad y de distribución al lanzar un dado



4.4.2 Variables aleatorias continuas

Concepto

Diremos que una variable aleatoria es continua cuando puede tomar cualquier valor en un intervalo. Por ejemplo, el peso de una persona, el tiempo de duración de un suceso, etc., corresponden a variables aleatorias continuas.

No es posible conocer el valor exacto de una variable continua, ya que medir su valor consiste en clasificarlo dentro de un intervalo: si el resultado de medir una longitud es 23 mm, todo lo que podemos afirmar es que la longitud real, no observable, está en el intervalo 22,5 mm a 23,5 mm. Los mo-

delos de variables aleatorias continuas se basan en este principio, y pueden caracterizarse mediante la función de densidad o la función de distribución.

Función de densidad

Supongamos, para concretar, que medimos una variable continua (longitud, tiempo, etc.) y representamos las medidas obtenidas en un histograma; es razonable admitir —y se ha comprobado repetidamente en la práctica— que, tomando más y más observaciones y haciendo clases cada vez más finas, el histograma tenderá a una curva suave que describirá el comportamiento a largo plazo de la variable estudiada. Llamaremos función de densidad a una función continua que verifica las condiciones

$$\text{(a) } f(x) \geq 0; \qquad \text{(b) } \int_{-\infty}^{\infty} f(x)dx = 1 \tag{4.12}$$

que puede interpretarse como la curva límite que obtendríamos en el histograma de una población disminuyendo indefinidamente las anchuras de cada clase.

El conocimiento de la función de densidad $f(x)$ permite calcular cualquier probabilidad por integración. Por ejemplo, la probabilidad de que la variable x sea menor que x_0 corresponde a sumar las probabilidades de todas las clases que contienen valores menores o iguales a x_0. Este resultado se obtiene fácilmente calculando el área bajo la función de densidad hasta el punto x_0 mediante:

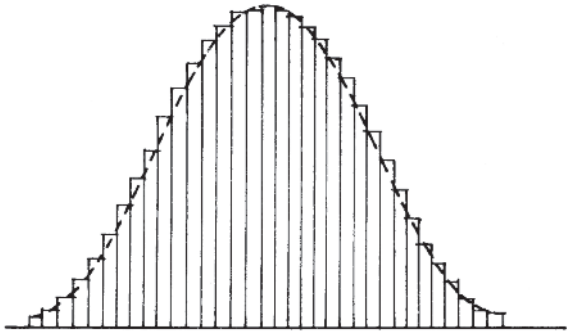
$$P(x \leqslant x_0) = \int_{-\infty}^{x_0} f(x)dx \tag{4.13}$$

Análogamente, la probabilidad de que la variable x tome un valor entre x_0 y x_1 se calculará como:

$$P(x_0 < x \leqslant x_1) = \int_{x_0}^{x_1} f(x)dx \tag{4.14}$$

La probabilidad de observar un valor cualquiera depende de la precisión con la que dicho valor se ha medido. Por ejemplo, la probabilidad al medir una longitud de observar el valor 12 cm es la probabilidad de que el verdadero valor esté entre 115 mm y 125 mm.

Figura 4.3 Histograma y función de densidad



En consecuencia, la probabilidad que un modelo de variable continua asigna a la observación de un valor exacto cualquiera (es decir, medido con infinita precisión) es cero. Esto es razonable, porque la frecuencia relativa de aparición de un número como 12,3401023297... puede considerarse cero si suponemos un número suficiente de cifras detrás de la coma. En contrapartida, la probabilidad de cualquier intervalo, por pequeño que sea, vendrá dada por el área que $f(x)$ encierra en ese intervalo. Si la base, Δx, es suficientemente pequeña, dicha área se aproxima por el área de un rectángulo de altura $f(x_0)$, siendo x_0 el centro del intervalo de longitud Δ_x, es decir:

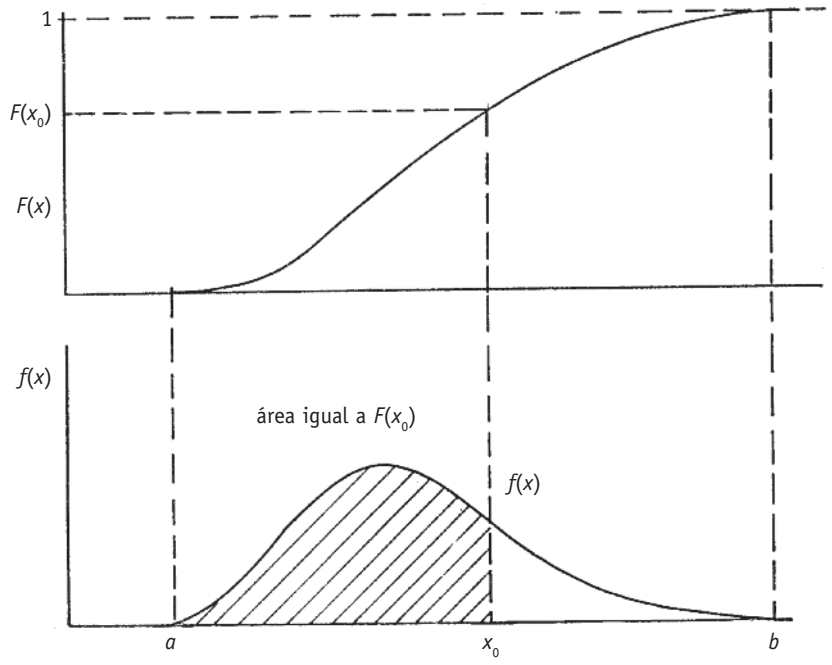
$$P(x_0 - \Delta x/2 < x \leqslant x_0 + \Delta x/2) \simeq f(x_0)\Delta x \qquad (4.15)$$

Una implicación de este resultado es que podemos olvidarnos del signo igual en las ecuaciones (4.13) y (4.14), ya que, para variables continuas,

$$P(a < x < b) = P(a \leq x < b) = P(a < x \leq b) = P(a \leq x \leq b)$$

En resumen, la función de densidad de probabilidad representa una aproximación muy útil para calcular probabilidades partiendo de un histograma: en primer lugar es mucho más simple, permite sustituir la tabla completa de valores de la distribución de frecuencias por la ecuación matemática de $f(x)$; en segundo lugar, es más general, trata de reflejar no el comportamiento de una muestra concreta, sino la estructura de distribución de los valores de la variable a largo plazo; en tercer lugar, es más operativa, permite obtener probabilidades de cualquier suceso. Esta tercera propiedad es clave: si disponemos del histograma de la distribución de unos datos medidos en metros, no es claro cómo calcular la probabilidad de que la variable esté en un intervalo (a, b) de amplitud 1 cm. Sin embargo, con la función de densidad esta pregunta tiene una respuesta inmediata: es el área encerrada por la función de densidad en dicho intervalo (a, b).

Figura 4.4 Función de densidad y distribución para una variable continua



Función de distribución

La función de distribución para una variable aleatoria continua se define como en el caso discreto por:

$$F(x_0) = P(x \leq x_0)$$

y teniendo en cuenta (4.13):

$$\boxed{F(x_0) = \int_{-\infty}^{x_0} f(x)dx} \qquad (4.16)$$

Así como en el caso discreto las diferencias entre dos valores consecutivos distintos de $F(x)$ proporcionan la función de probabilidad, para variables continuas la derivada de $F(x)$ proporciona la función de densidad. En efecto, utilizando (4.15):

$$F(x_0 + \Delta x) - F(x_0) = P(x_0 < x \leq x_0 + \Delta x) \simeq f(x_0)\Delta x$$

con lo que concluimos que:

$$\boxed{f(x) = \frac{dF(x)}{dx}} \tag{4.17}$$

La función de distribución de una variable continua será una función continua que verifica las tres propiedades básicas estudiadas para variables discretas:

1) $F(-\infty) = 0$, ya que $\int_{-\infty}^{-\infty} f(x)dx = 0$.

2) $F(+\infty) = 1$, ya que, por construcción, $\int_{-\infty}^{\infty} f(x)dx = 1$.

3) Es no decreciente: si $x_1 > x_2$, $F(x_1) \geq F(x_2)$.

La figura 4.4 ilustra la relación entre la función de densidad y distribución. En un punto x_0, la función de distribución indica la probabilidad de que la variable sea menor o igual que x_0, que es el área rayada en la función de densidad. La ordenada $f(x_0)$ en ese punto no es una probabilidad, aunque si la multiplicamos por la longitud de un intervalo pequeño, Δx, obtenemos la probabilidad de que la variable se encuentre en dicho intervalo.

Ejemplo 4.8

Se dice que una variable aleatoria continua es *uniforme* en un intervalo (a, b) si su función de densidad es constante en dicho intervalo y nula fuera de él. Calcular la función de densidad y de distribución para una variable uniforme (0,10) y la probabilidad de que la variable esté en el intervalo (1; 1,5), contenido dentro del (a, b).

Si $f(x) = k$, como:

$$\int_a^b f(x)dx = 1 = k(b - a) = 1$$

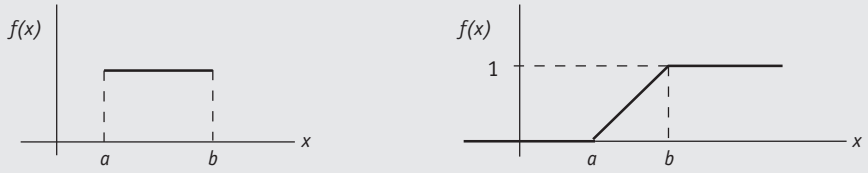
resulta $f(x) = (b - a)^{-1}$. La función de distribución será:

$$F(x) = \int_a^x \frac{dx}{b - a} = \frac{x - a}{b - a}$$

La figura 4.5 presenta estas funciones. La probabilidad pedida es:

$$P(1 < x < 1{,}5) = \int_{1}^{1{,}5} \frac{1}{b-a}\,dx = \frac{0{,}5}{b-a}$$

Figura 4.5 Función de densidad y distribución de una variable uniforme (*a*, *b*)



4.4.3 Medidas características de una variable aleatoria

Podemos construir medidas características de la distribución de una variable aleatoria análogamente a como lo hicimos para una distribución de frecuencias en el capítulo 2. Es costumbre representar estas medidas teóricas por letras griegas, para diferenciarlas de las calculadas sobre datos, que se representan con letras romanas.

Medidas de centralización

La medida de centralización más utilizada es la *media* (μ) o *esperanza matemática, E*(x), de la variable, que se obtiene promediando cada posible valor por su probabilidad. En el caso discreto:

$$\mu = E(x) = \Sigma x_i p(x_i) \tag{4.18}$$

donde el sumatorio va extendido a todos los valores posibles de la variable. En el caso continuo esta fórmula se convierte en:

$$\mu = E(x) = \int_{-\infty}^{\infty} xf(x)dx \tag{4.19}$$

La segunda medida importante de centralización es la mediana que, en términos intuitivos, es aquel valor que divide la probabilidad total en dos

partes iguales. Para una variable continua la mediana será pues un valor m definido por:

$$F(m) = 0{,}5 = P(x \leq m)$$

Para variables discretas definiremos la *mediana* como el menor valor de la variable que satisface

$$F(x) \geqslant 0{,}5$$

Finalmente, la *moda* es el valor más probable.

Ejemplo 4.9

Calcular el beneficio esperado (o beneficio medio) con una apuesta de 100 euros a la ruleta: (a) a un número cualquiera; (b) a rojo frente a negro.

Los resultados posibles de una jugada en la ruleta son los números (0, 1, ..., 36) con probabilidades 1/37. Si apostamos 100 euros a un número, la variable aleatoria x, beneficio obtenido, tomará los valores siguientes:

$x = -100$, si ocurre cualquier número distinto al apostado, $P(x = -100) = 36/37$.
$x = 3.500$, si ocurre el número elegido, $P(x = 3.500) = 1/37$

Por tanto:

$$E(x) = -100\ (36/37) + 3.500\ (1/37) = -2{,}7 \text{ euros}$$

que supone una pérdida del 2,7% de la cantidad invertida.

En el segundo caso hay dos resultados posibles: +100 (si sale rojo), −100 (si sale negro o el cero). Entonces:

$$E(x) = -100\ (19/37) + 100\ (18/37) = -2{,}7 \text{ euros}$$

que es el mismo resultado anterior. Todas las apuestas de la ruleta tienen la misma esperanza de pérdida.

Ejemplo 4.10

Calcular la esperanza de beneficios para una compañía de seguros al hacer un seguro cuya prima anual es r, la probabilidad de siniestro p y la cantidad asegurada M.

Los resultados posibles para la compañía son: con probabilidad $(1 - p)$ no siniestro y gana r; con probabilidad p ocurre el siniestro y pierde $M - r$, ya que en cualquier caso cobra la prima. Por tanto:

$$E(x) = (1 - p)r - p(M - r) = r - pM$$

Por tanto, si $r > pM$ el beneficio a largo plazo está asegurado si efectúa un gran número de seguros de este tipo. Por ejemplo, si $r = 200$ euros, $p = 0{,}001$ y $M = 100.000$, el beneficio esperado es de 100 euros [200 – – (0,001) (100.000)] por asegurado.

Medidas de dispersión

Como en las distribuciones de frecuencias, podemos asociar a cada medida de centralización una de dispersión. A la media se le asocia la *desviación típica,* cuyo cuadrado es la *varianza*, definida para variables continuas por:

$$Var\,(x) = \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx \qquad (4.20)$$

Para variables discretas las integrales se convierten en sumas y las probabilidades $p(x)$ sustituyen a los elementos de probabilidad $f(x)dx$.

El *percentil p* de una variable aleatoria x es el valor x_p que verifica

$$p(x < x_p) \leq p$$
$$p(x \leq x_p) \geq p$$

Para variables continuas las dos condiciones anteriores equivalen a:

$$F(x_p) = p$$

Los *cuartiles* dividen la distribución en cuatro partes iguales. La mediana coincide con el segundo cuartil y con el percentil 0,5. La medida absoluta de dispersión más utilizada es el *rango intercuartílico,* que es la diferencia entre el tercer y primer cuartil $(Q_3 - Q_1)$ y representa la zona central donde se encuentra el 50% de la probabilidad.

$$\text{Rango intercuartílico} = Q_3 - Q_1 \qquad (4.21)$$

Para distribuciones simétricas $Q_2 - Q_1 = Q_3 - Q_2$ y, por tanto, el rango intercuartílico es el doble de la distancia entre la mediana y los cuartiles.

La medida de dispersión que se asocia a la mediana es la Meda, que es la mediana de las distancias en valor absoluto entre la variable y la mediana:

$$\text{Meda} = \text{Mediana}\,(|x - \text{Med}(x)|) \tag{4.22}$$

Para distribuciones simétricas, el 50% de las desviaciones son menores que $(Q_3 - \text{Med}) = (\text{Med} - Q_1)$ y el 50% mayores. En efecto, son menores las de todos los valores x que verifican $Q_1 \leq x \leq Q_3$ y mayores las de los puntos no incluidos en el intervalo $(Q_1 Q_3)$. En consecuencia, la Meda será $(Q_3 - \text{Med})$, es decir, la mitad del rango intercuartílico. Este resultado sólo es cierto para distribuciones simétricas.

Otras medidas características

En general, definimos *momento de orden k respecto al origen*, m_k, de una variable aleatoria continua por:

$$m_k = \int x^k f(x)dx \tag{4.23}$$

y *momento de orden k respecto a la media* por:

$$\mu_k = \int (x - \mu)^k f(x)dx \tag{4.24}$$

El *coeficiente de asimetría* se define por:

$$CA = \frac{\mu_3}{\sigma^3} \tag{4.25}$$

el de *apuntamiento* por:

$$CAp = \frac{\mu_4}{\sigma^4} \tag{4.26}$$

y el de *variación* por:

$$CV = \frac{\sigma}{|\mu|} \tag{4.27}$$

En el caso de variables aleatorias discretas, las integrales se convierten en sumas, y las probabilidades $p(x)$ sustituyen a los elementos de *probabilidad* $f(x)dx$. La interpretación de estos coeficientes para variables aleatorias es idéntica a la expuesta en el capítulo 2 para distribuciones de frecuencias.

Acotación de Tchebychev

Conocer la media y la desviación típica de una variable aleatoria discreta o continua permite calcular la proporción de la distribución que está situada entre $\mu \pm k\sigma$, siendo k una constante positiva. Se verifica que:

$$P(\mu - k\sigma \leqslant x \leqslant \mu + k\sigma) \geqslant 1 - 1/k^2 \tag{4.28}$$

para cualquier valor de k. Esta propiedad se demostró para distribuciones de frecuencias (en el capítulo 2) y su generalización para variables aleatorias discretas o continuas es inmediata.

La fórmula (4.28) indica que, para cualquier variable aleatoria, el intervalo $\mu \pm 3\sigma$ ($k = 3$) contiene, al menos, el 89% de la distribución y el $\mu \pm \pm 4\sigma$ ($k = 4$) el 94%.

4.4.4 Transformaciones

Interesa con frecuencia obtener la distribución de una función conocida de una variable aleatoria. Por ejemplo, queremos analizar los datos en logaritmos para obtener una distribución más simétrica o cambiar la escala de medida de la variable (metros por centímetros o dólares por pesetas, por ejemplo). En general, dada una variable aleatoria x, se desea obtener la distribución de otra variable $y = h(x)$ donde la función h es conocida.

Función de distribución

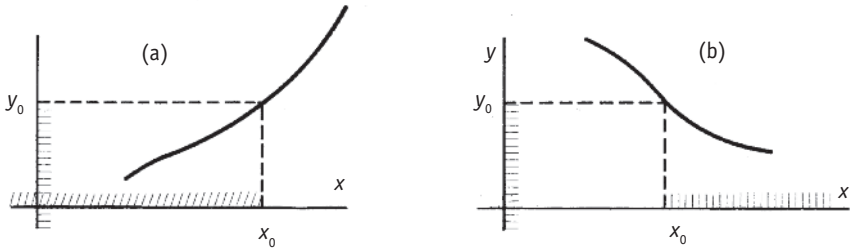
Llamando $G(y)$ a la función de distribución de la nueva variable aleatoria $y = h(x)$, tendremos:

$$G(y_0) = P(y \leq y_0) = P[h(x) \leq y_0] = p(x \in A) \tag{4.29}$$

donde A representa el conjunto de valores de x en los que se verifica que $h(x)$ es menor o igual a y_0.

En el caso particular de variables continuas en que la función h sea continua y monótona creciente (figura 4.6a), la relación $h(x) \leq y_0$ equivale a $x \leq \leq x_0$, donde $y_0 = h(x_0)$ o bien $x_0 = h^{-1}(y_0)$ y podemos escribir:

Figura 4.6 Relación entre intervalos de *y* y de *x* para funciones monótonas



$$G(y_0) = P[x \le h^{-1}(y_0)] = F[h^{-1}(y_0)] = F(x_0) \tag{4.30}$$

donde F es la función de distribución de la variable x.

Si h es monótona decreciente (figura 4.6b), la relación $y \le y_0$ equivale a la $x \ge x_0$, donde $y_0 = h(x_0)$ y tendremos:

$$G(y_0) = P(x \ge x_0) = 1 - P(x \le x_0) = 1 - F(x_0) \tag{4.31}$$

Función de probabilidad y de densidad

Para variables discretas la función de probabilidad de y será:

$$p(y_0) = P(y = y_0) = \sum_{y_0 = h(x_i)} p(x_i)$$

es decir, para calcular la probabilidad de y_0 sumamos las probabilidades de todos los valores de x que dan lugar a y_0.

Para variables continuas derivaremos en la función de distribución para obtener la función de densidad. En el caso (4.30), aplicando la regla de la cadena y llamando $g(y)$ a la función de densidad:

$$\frac{dG(y)}{dy} = g(y) = \frac{dF(x)}{dx} \cdot \frac{dx}{dy} = f(x) \frac{dx}{dy} \tag{4.32}$$

Observemos que, como h es creciente, $dx|dy$ es positiva y $g(y)$ es siempre positiva, como corresponde a una función de densidad.

En el caso (4.31), procediendo análogamente:

$$g(y) = f(x) \frac{dx}{dy} \tag{4.33}$$

y como ahora $dx|dy$ es negativa, ambos casos pueden escribirse de forma unificada en:

$$g(y) = f(x) \left| \frac{dx}{dy} \right| \tag{4.34}$$

Para interpretar este resultado observemos que la función de densidad *tiene unidades:* probabilidad por unidad de medida de la variable. En consecuencia, al cambiar estas unidades de medida, la función de densidad tendrá que variar correspondientemente. Supongamos, por ejemplo, que $g_m(y)$ representa la función de densidad de una longitud medida en metros y $f_{cm}(x)$ representa la función de densidad de la misma longitud medida en centímetros. La probabilidad de que la longitud sea igual a 0,2 cm (entendiendo por ello que esté entre 1,5 y 2,5 mm) deberá ser la misma usando cualquiera de las dos funciones. Como:

$$P(y = 0{,}002 \text{ m}) = P(0{,}0015 < y < 0{,}0025) = g_m(0{,}002)\ 0{,}001$$
$$P(x = 0{,}2 \text{ cm}) = P(0{,}15 < x < 0{,}25) = f_{cm}(0{,}20)\ 0{,}1$$

para que ambas probabilidades coincidan:

$$g_m(0{,}002) = f_{cm}(0{,}20)\ 100$$

que expresa que la función de densidad en metros para un valor y se obtiene multiplicando el valor de la función de densidad para dicho valor en cm (x) por el ratio de las longitudes de los intervalos unidad. Es decir:

$$g(y) = f(x) \left| \frac{\Delta x}{\Delta y} \right|$$

En general, si hemos medido la variable x en una escala donde la función de densidad es $f(x)$, y pasamos a otra variable, y, mediante una transformación biunívoca cualquiera:

$$y = h(x)$$

la función de densidad de y deberá verificar que la probabilidad asignada a un intervalo Δy_0, con centro y_0, sea la misma que la asignada al intervalo correspondiente de x, Δx_0, con centro x_0, siendo $y_0 = h(x_0)$. Por tanto:

$$g(y_0)|\Delta y_0| = f(x_0)|\Delta x_0|$$

y en el límite, eliminando subíndices cuando $\Delta x \to 0$, $\Delta y \to 0$, se obtiene la fórmula (4.34), que es la ecuación básica de cambio de variable para funciones de densidad. Especifica que, para obtener la función de densidad de una variable aleatoria que es una función biunívoca de otra conocida (x), basta sustituir en la función de densidad conocida la variable x por su expresión en función de y y multiplicar por la derivada, que representa el cambio de escala inducido por la transformación.

El apéndice 4B generaliza este resultado para funciones no biunívocas.

Esperanzas

La media o esperanza matemática de una variable, que es función de otra con distribución conocida, puede calcularse directamente sin necesidad de obtener la nueva distribución. Vamos a demostrar que si $y = h(x)$

$$\boxed{E[h(x)] = \int_{-\infty}^{\infty} h(x)f(x)dx} \tag{4.35}$$

donde si x es una variable discreta, la integral se convierte en suma y la función de probabilidad $p(x)$ reemplaza a $f(x)dx$.

Este resultado es totalmente general, pero lo demostraremos para el caso particular en que x es continua y $h(x)$ es continua y monótona creciente. Entonces, utilizando (4.32):

$$E[y] = \int_{-\infty}^{\infty} yg(y)dy = \int_{-\infty}^{\infty} h(x)f(x)dx = E[h(x)]$$

La ecuación (4.35) justifica escribir la varianza de una variable aleatoria como:

$$\sigma^2 = E(x - \mu)^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx$$

En efecto, $(x - \mu)^2$ será otra variable aleatoria, y su esperanza, según (4.35), es la varianza de la variable.

En particular, si $y = a + bx$:

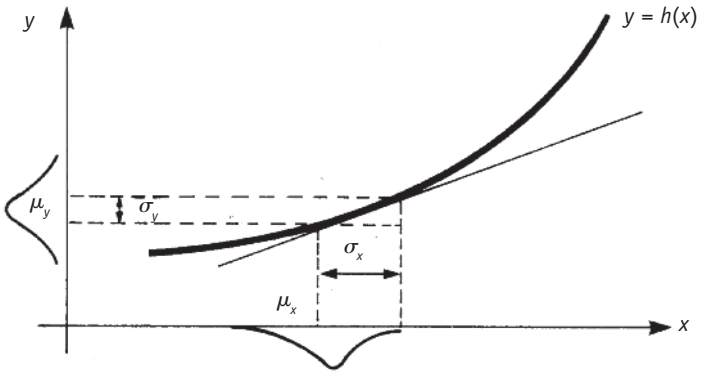
$$E[y] = E[a + bx] = a + bE[x]$$

ya que, aplicando de nuevo (4.35), la esperanza de una constante es ella misma.

Análogamente se obtiene:

$$\text{Var}[y] = E(y - \mu_y)^2 = E[b(x - \mu_x)]^2 = b^2 \text{ Var}(x)$$

Figura 4.7 Modificación de la variabilidad por una transformación



es decir, si efectuamos una transformación lineal, la varianza de la nueva variable es un múltiplo de la original.

Cuando la transformación realizada es no lineal pero continua y con derivadas continuas, podemos obtener una expresión aproximada de los momentos de y en función de los de x desarrollando la función en serie de Taylor:

$$y = h(x) \simeq h(\mu_x) + (x - \mu_x)h'(\mu_x) + \frac{1}{2}(x - \mu_x)^2 h''(\mu_x) \qquad (4.36)$$

y tomando esperanzas en ambos miembros:

$$\boxed{E[y] \simeq h(\mu_x) + \sigma_x^2 h''(\mu_x)/2} \qquad (4.37)$$

Si despreciamos términos de segundo grado y aproximamos $E[y]$ por $h(\mu_x)$, la figura 4.7 muestra que, aproximadamente:

$$\boxed{\sigma_y \simeq \sigma_x |h'(\mu_x)|} \qquad (4.38)$$

Esta expresión también se obtiene despreciando el término cuadrático en (4.36) y escribiendo

$$y - E(y) \simeq (x - \mu_x)h'(\mu_x)$$

y elevando al cuadrado y tomando esperanzas se obtiene (4.38).

Ejemplo 4.11

La variable y tiene función de densidad $f(y) = e^{-y}$ para $y > 0$, $f(y) = 0$ en otro caso. Se pide: 1) su función de distribución; 2) su esperanza; 3) su mediana; 4) el percentil 0,9; 5) la función de densidad de una variable x relacionada con la anterior por:

$$y = (x/b)^c$$

siendo b y c constantes positivas; 6) la moda de la distribución de x.

1) Para calcular $F(y)$ aplicaremos la definición:

$$F(y_0) = \int_{-\infty}^{y_0} f(y)dy = \begin{cases} 0 \text{ para } y_0 \leq 0 \\ \int_0^{y_0} e^{-y}dy = 1 - e^{-y_0} \text{ para } y_0 > 0 \end{cases}$$

Comprobaremos que la función así construida es una función de distribución:

a) $F(-\infty) = 0$, ya que F es cero para cualquier valor negativo.
b) $F(+\infty) = 1 - e^{-\infty} = 1$.
c) Si $y_1 > y_2$, $F(y_1) > F(y_2)$, ya que $e^{-y_1} < e^{-y_2}$

2) La esperanza de la variable será:

$$E[y] = \int_0^x ye^{-y}dy = 1$$

(véase apéndice 5A para la integración por partes de esa función).

3) La mediana se obtendrá por:

$$0{,}5 = 1 - e^{-\text{Med}}$$

$$\text{Med} = -\ln 0{,}5 = 0{,}69$$

y el 50% de la distribución está por debajo de 0,69. La distribución es muy asimétrica, por debajo de la media, como:

$$F(y) = 1 - e^{-1} = 0{,}63$$

se encuentra el 63% de la distribución.

4) Para encontrar el percentil 0,9, que llamaremos $y_{0,9}$, definido por:

$$P(y \leqslant y_{0,9}) = F(y_{0,9}) = 0{,}9$$

utilizando la expresión de la función de distribución,

$$0{,}9 = 1 - e^{-y_{0,9}}$$

$$y_{0,9} = -\ln 0{,}1 = 2.30$$

5) La función de densidad de x será:

$$f(x) = 0 \qquad x < 0$$

$$f(x) = f(y) \left| \frac{dy}{dx} \right| \qquad x > 0$$

como:

$$\frac{dy}{dx} = \frac{c}{b} \left(\frac{x}{b} \right)^{c-1}$$

Se obtiene:

$$f(x) = \frac{c}{b} \left(\frac{x}{b} \right)^{c-1} \exp \{-(x/b)^c\}$$

Ésta es la distribución de Weibull que se utiliza para el estudio de duraciones de vida de materiales y fiabilidad de componentes.

6) Para calcular la moda, utilizaremos que $f(x)$ y $\ln f(x)$ tienen el mismo máximo y derivaremos en el logaritmo de la función de densidad. Entonces:

$$\ln f(x) = \ln \frac{c}{b} + (c - 1) \ln \frac{x}{b} - \left(\frac{x}{b} \right)^c$$

$$\frac{d \ln f(x)}{dx} = 0 = (c - 1) \frac{1/b}{x/b} - c \left(\frac{x}{b} \right)^{c-1} \frac{1}{b}$$

$$x = [(c - 1)/c]^{1/c} b \qquad \text{para } c > 1.$$

Ejercicios 4.2

4.2.1. Representar la función de distribución para la variable aleatoria suma de las caras al tirar un dado dos veces.

4.2.2. Obtener la constante k para que la función $f(x) = k$ represente la función de densidad de una variable continua en el intervalo (a, b).

4.2.3. Dibujar la función de distribución del problema anterior y calcular con ella la mediana de la distribución.

4.2.4. Dada la variable aleatoria x con función de distribución

$$F(x) = \begin{cases} 0 & x \leq 0 \\ x^n & 0 < x \leq 1 \\ 1 & x > 1 \end{cases}$$

donde $n \geq 1$ se pide:

a) Calcular la función de densidad.
b) Encontrar la mediana.
c) Encontrar la media y la varianza de x.

4.2.5. Una variable aleatoria tiene como función de densidad:

$$f(x) = \begin{cases} 3x^2 & 0 < x < 1 \\ 0 & \text{en otro caso} \end{cases}$$

Se pide:

a) La función de distribución de x.
b) Encontrar $F(2/3)$, $F(9/10)$ y $P(1/3 < x \leq 1/2)$.
c) Aquel valor de a tal que $P(x \leq a) = 1/4$.
d) La media y varianza de x.

4.2.6. Una máquina fabrica ejes cuyos radios se distribuyen con función de densidad $f(x) = k(x - 1)(3 - x)$ si $1 \leq x \leq 3$ y cero en otro caso. La variable x se mide en m. Se pide:

a) Calcular k.
b) Escribir la función de densidad para los radios de los ejes medidos en cm.
c) Escribir la función de densidad para el diámetro de los ejes.
d) Escribir la función de densidad para el área de las secciones.
e) Si los ejes se desechan cuando su radio se desvía de *2* metros más de 80 cm, calcular la proporción de ejes que serán rechazados.

4.2.7. Calcular el valor esperado de una apuesta a blanco o negro a la ruleta.

4.2.8. La probabilidad de un tipo de accidente industrial en un año es 0,001. Una compañía de seguros propone a una empresa un seguro de accidentes cuyo coste anual es de 10.000 pesetas, comprometiéndose en caso de accidente a satisfacer una cantidad de 5 millones de pesetas en concepto de indemnización. Calcular el beneficio esperado para la compañía de seguros.

4.2.9. Dada la función de distribución $F(x) = (x - a)/(b - a)$, obtener la función de densidad y dibujarla.

4.2.10. El tiempo de reparar una máquina en horas tiene la función de distribución: $F(x) = 0$, $x \leq 0$; $F(x) = x/2$, $0 \leq x \leq 1$; $F(x) = 1/2$, $1 \leq x \leq 2$; $F(x) = x/4$, $2 \leq x \leq 4$; $F(x) = 1$, $x \geq 4$.
   a) Dibujar la función de distribución.
   b) Obtener la función de densidad e interpretarla.
   c) Si el tiempo de reparación es superior a 1 hora, ¿cuál es la probabilidad de que sea superior a 3,5 horas?

4.2.11. Si x tiene $f(x) = 1$ $(0 < x \leq 1)$, calcular la función de densidad de las variables:
   a) $y = x^2$.
   b) $y = \sqrt{x}$.
   c) $y = 1/x$.

4.2.12. La variable aleatoria x tiene la siguiente función de densidad: para $0 < x < 2$, $f(x) = mx$; para $2 < x < 4$, $f(x) = 1 - mx$. Se pide:
   a) Hallar m.
   b) Hallar $E(x)$.
   c) Dibujar $F(x)$.

4.2.13. Los ejes del problema 4.2.6 pueden acoplarse entre sí siempre que sus radios estén entre 1,7 y 2,4 metros. Tomamos cinco ejes al azar. ¿Cuál es la probabilidad de que puedan acoplarse entre sí?

4.5 Resumen del capítulo

Este capítulo presenta las reglas básicas de construcción de modelos para el tipo de datos considerado en el capítulo 2. La probabilidad es un modelo para las frecuencias relativas y además un procedimiento general para cuantificar la incertidumbre. Una variable aleatoria es un modelo para una variable observable cuyo valor no se conoce a priori. Las variables aleatorias se clasifican, de la misma forma que las variables observables estudia-

das en el capítulo 2, como discretas o continuas. Todas las variables aleatorias quedan definidas por la función de distribución, pero habitualmente las variables discretas se definen por su función de probabilidades y las continuas por su función de densidad. De la misma forma que definimos medidas de centralización, dispersión, asimetría y curtosis o apuntamiento para los datos reales, podemos definir estas medidas para variables aleatorias con una interpretación análoga.

Cuando transformamos una variable aleatoria discreta es muy simple obtener la nueva función de probabilidad de la nueva variable: sólo tenemos que sumar las probabilidades de los valores de la primera que conducen al mismo valor de la segunda. Al transformar una variable aleatoria continua para obtener la función de densidad, es importante tener en cuenta que la función de densidad proporciona probabilidad por unidad de longitud, y habrá que ajustar la densidad de la nueva variable por el cambio en las longitudes que produce la transformación. La media y varianza de la variable transformada pueden calcularse aproximadamente de forma rápida a partir de la media y la varianza de la variable original.

4.6 Lecturas recomendadas

Un libro clásico lleno de ilustraciones sobre la aplicación de la probabilidad es Feller (1971). En español, Cramer (1968), Parzen (1987), Castillo (1978) y Quesada y García (1988) son referencias adecuadas, las primeras más simples y las dos últimas con mayor rigor matemático. En inglés, Gnedenko (1998), Papoulis (2002), Trivedi (2002) y Ross (2005) son referencias clásicas.

La concepción subjetiva de la probabilidad ha sido claramente expuesta por De Finetti (1974) en una obra importante y muy pedagógica. Lindley (1970) presenta una introducción más sucinta; O'Hagan (1988), un tratamiento muy claro y actual; y Bernardo y Smith (2000), un estudio profundo y documentado de este enfoque.

La fundamentación matemática de la probabilidad como una parte de la teoría de la medida se encuentra en Kolmogorov (1956); un tratado riguroso en esta línea es Loeve (1976).

Todos los manuales de estadística referenciados en la bibliografía incluyen una parte de cálculo de probabilidades. Especialmente claros son Guttman *et al.* (1982) y Larsen y Max (2005).

Apéndice 4A: Álgebras de probabilidad

Experimentos. Sucesos

Definiremos un experimento como un proceso de observación de la realidad que puede repetirse en condiciones idénticas. Ejemplos posibles son:

a) Tirar un dado y observar el número que sale en la cara superior.
b) Contar el número de clientes que llegan a un puesto de servicio en un día.
c) Medir el tiempo que transcurre en una centralita entre dos llamadas.

Definir un experimento requiere:

a) Especificar el conjunto de condiciones en que se realiza.
b) Indicar cuáles son los resultados posibles.

Es conveniente definir el conjunto de resultados con toda generalidad. Por ejemplo, en el experimento (b) tomaremos como conjunto de resultados el de los números naturales, con lo que:

a) Describimos lo que *puede* en principio ocurrir independientemente de sus posibilidades de ocurrencia, que son aspectos del fenómeno conceptualmente diferentes.
b) No introducimos restricciones al número de clientes posibles.

En el ejemplo (c), por las razones anteriores, un resultado será cualquier valor del eje positivo real.

Sea cual sea la naturaleza de los resultados, los representaremos por letras *a, b, c...* y los llamaremos sucesos elementales. El número de sucesos elementales posibles puede ser finito [como en el ejemplo (a)], infinito pero numerable [ejemplo (b)] o infinito [ejemplo (c)]. El conjunto de todos los resultados elementales lo representaremos por E y lo llamaremos *espacio muestral* o espacio de todos los sucesos elementales.

Llamaremos suceso, y lo representaremos por letras *A, B, C...,* a un subconjunto de sucesos elementales.

Diremos que el suceso A ha ocurrido cuando el resultado del experimento es un suceso elemental contenido en A. Por lo tanto, el espacio muestral es un suceso que siempre ocurre, y le llamaremos suceso seguro.

Sea $\mathcal{A}$ una familia de sucesos que contiene a E, suceso seguro, y además un número cualquiera de elementos. Entre los elementos del conjunto $\mathcal{A}$ definimos la unión (que representaremos por +), intersección (que representaremos por ·), complemento (que representaremos para un conjunto A por $\bar{A}$) y diferencia (que escribiremos –) de sucesos de la forma habitual.

Exigiremos que operando con dichas leyes de composición:

a) $\begin{matrix} \text{Si } A \in \mathcal{A} \\ B \in \mathcal{A} \end{matrix}$ y $C = A + B \Rightarrow C \in \mathcal{A}$

b) $\begin{matrix} \text{Si } A \in \mathcal{A} \\ B \in \mathcal{A} \end{matrix}$ y $D = A \cdot B \Rightarrow D \in \mathcal{A}$

c) Si $A \in \mathcal{A} \Rightarrow \overline{A} \in \mathcal{A}$

En estas condiciones, la clase $\mathcal{A}$ es cerrada para estas operaciones y se denomina álgebra. Dado que contiene a E, contiene a su complementario, el conjunto vacío $\varnothing$, que llamaremos suceso imposible. Es conveniente desde un punto de vista matemático trabajar con una clase que sea cerrada no sólo para la unión finita, sino también para las uniones e intersecciones numerables de sucesos. Llamaremos σ-álgebra a una clase $\mathcal{A}$ de subconjuntos de E que tiene las propiedades:

1) $\text{Si } A_i \in \mathcal{A} \Rightarrow \forall_i \sum_{i=1}^{\infty} A_i \in \mathcal{A}$

2) $\text{Si } A_i \in \mathcal{A} \Rightarrow \forall_i \prod_{i=1}^{\infty} A_i \in \mathcal{A}$

3) $\text{Si } A \in \mathcal{A} \Rightarrow \overline{A} = E - A \in \mathcal{A}$

De esta manera garantizamos que con las operaciones entre conjuntos establecidas en $\mathcal{A}$ obtendremos siempre a partir de elementos de $\mathcal{A}$ nuevos elementos de este conjunto.

La restricción de considerar clases cerradas de sucesos es lógica y conviene desde un punto de vista matemático. Dados dos sucesos A y B, es razonable preguntarse por el suceso que ocurra o bien A o bien B (suceso $A + B$) o que ocurran simultáneamente ambos (suceso $A \cdot B$). La extensión a una σ-álgebra admitiendo las uniones e intersecciones numerables es conveniente para trabajar con espacios maestrales generales.

Dados dos sucesos cualesquiera de $\mathcal{A}$, diremos que son mutuamente excluyentes o disjuntos si:

$$A \cdot B = \emptyset$$

Un conjunto de sucesos $A_1, \ldots, A_n$ es exhaustivo si:

$$\sum_{i=1}^{n} A_i = E$$

Si además los sucesos son todos disjuntos, es decir:

$$\forall ij,\ i \neq j \qquad A_i \cdot A_j = \emptyset$$

$$\sum_{i=1}^{n} A_i = E$$

diremos que constituyen una clase completa de sucesos o una partición del espacio muestral. Obviamente el conjunto de los sucesos elementales es siempre una clase completa de sucesos y representa la partición más fina del espacio muestral.

Probabilidad

Supongamos un experimento definido por un conjunto de condiciones $\mathscr{C}$, un espacio muestral E, una clase de sucesos $\mathscr{A}$ con estructura de σ-álgebra. Se trata de establecer una medida de incertidumbre para los sucesos de este experimento. Postularemos que esta medida debe tener las siguientes propiedades: es una función de conjunto que asocia a los sucesos (subconjuntos) de una clase de conjuntos un número real, tal que:

1.º) $\forall A\ P(A) \geq 0$
2.º) $P(E) = 1$
3.º) $A_i \in \mathscr{A} \quad \forall i, \qquad$ y $\qquad A_i \cdot A_j = \emptyset \quad \forall ij; \quad i \neq j$

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

La formalización anterior es debida a Kolmogorov (1933). La tripleta $(E, \mathscr{A}, P)$ se denomina un espacio de probabilidades. A partir de estos axiomas se demuestran fácilmente las propiedades siguientes:

1) $P(\bar{A}) = 1 - P(A)$
2) Si $A \subseteq B \Rightarrow P(A) \leq P(B)$
3) Si $\emptyset$ es el suceso imposible $P(\emptyset) = 0$; como $\bar{E} = \emptyset$, entonces:

$$P(\bar{E}) = P(\emptyset) = 1 - P(E) = 0$$

4) $\forall A \in \mathscr{A} \quad 0 \leq P(A) \leq 1$, ya que:

$$\emptyset \subseteq A \subseteq E$$

y por la propiedad 3) es inmediato.

5) Dados dos sucesos cualesquiera A y B no necesariamente excluyentes:

$$P(A + B) = P(A) + P(B) - P(A \cdot B)$$

6) Dados n sucesos cualesquiera $A_1, ..., A_n$:

$$P\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i=1}^{n} \sum_{j=i+1}^{n} P(A_i \cdot A_j) +$$

$$+ \sum_{i=1}^{n} \sum_{j=i+1}^{n} \sum_{k=j+1}^{n} P(A_i \cdot A_j \cdot A_k) - ... + (-1)^{n+1} P(A_1 \cdot A_2 ... A_n)$$

7) $P(\Sigma A_i) \leq \Sigma P(A_i)$ (desigualdad de Bonferroni).
8) $P(A \cdot B) \geq 1 - P(\overline{A}) - P(\overline{B})$ (desigualdad de Boole).

En el texto eliminaremos el · para indicar la intersección de sucesos.

Apéndice 4B: Cambio de variable en el caso general

Si la relación $y = h(x)$ que define la nueva variable no es biunívoca, tendremos que determinar todos los puntos que verifican $x = h^{-1}(y)$ y sumar los elementos de probabilidad correspondientes a todos ellos. Sean: $x_1, ..., x_i, ... x_n$ estos puntos. Entonces:

$$f(y) = \sum_{i=1}^{n} f(x_i) \left| \frac{dx_i}{dy} \right|$$

Por ejemplo, si $y = x^2$ y la variable x toma valores negativos, la relación no es biunívoca, $x_1 = \sqrt{y}$; $x_2 = -\sqrt{y}$:

$$f(y) = f_x(\sqrt{y}) \frac{1}{2\sqrt{y}} + f_x(-\sqrt{y}) \frac{1}{2\sqrt{y}}$$

5. Modelos univariantes de distribución de probabilidad

Carl Friederich Gauss (1777-1855)
Matemático alemán de extraordinaria precocidad. Inventa a los 18 años el método de mínimos cuadrados y propone la distribución normal para representar los errores de observación. Fue director del observatorio astronómico de Göttingen y considerado el mejor matemático de su tiempo. Sus contribuciones a la astronomía y las matemáticas son enormes.

5.1 El proceso de Bernoulli y sus distribuciones asociadas

5.1.1 Proceso de Bernoulli

Supongamos un experimento donde se observan elementos de una población, con las siguientes características:

1) La observación consiste en clasificarlos en dos categorías, que llamaremos A (aceptable) y D (defectuoso).
2) La proporción de elementos A y D en la población es constante y no se modifica cualquiera que sea la cantidad observada. Esto implica que si la población es finita, los elementos se reemplazan una vez observados. Llamaremos p a la probabilidad de defectuoso, y $q = 1 - p$ a la de aceptable.

3) Las observaciones son independientes, es decir, la probabilidad de elemento defectuoso es siempre la misma y no se modifica por cualquier combinación de elementos defectuosos o aceptables observados.

Este modelo se aplica a poblaciones finitas de las que tomamos elementos al azar con reemplazamiento, y también a poblaciones conceptualmente infinitas, como las piezas que producirá una máquina, siempre que el proceso generador sea estable (proporción de piezas defectuosas constante a largo plazo) y sin memoria (el resultado en cada momento es independiente de lo previamente ocurrido).

Ejemplos de procesos de Bernoulli son observar el sexo de un recién nacido, si un cliente está satisfecho o no con un servicio, la aparición del número 10 en tiradas sucesivas de una ruleta o la aparición de un elemento defectuoso en una fabricación.

En este proceso podemos definir distintas variables aleatorias que darán lugar a distintas distribuciones de probabilidad.

5.1.2 Distribución de Bernoulli

Definimos la variable aleatoria de Bernoulli por:

$$x = \begin{cases} 0 \text{ si el elemento es aceptable} \\ 1 \text{ si el elemento es defectuoso} \end{cases}$$

La función de probabilidades de esta variable se escribe:

$$\boxed{P(x) = p^x\, q^{1-x}; \qquad x = 0,1} \tag{5.1}$$

Su media será:

$$\mu = E(x) = 0 \cdot (1 - p) + 1 \cdot p = p \tag{5.2}$$

y la desviación típica:

$$DT(x) = \sigma = \sqrt{(0 - p)^2(1 - p) + (1 - p)^2 p} = \sqrt{pq} \tag{5.3}$$

En esta distribución la media y la variabilidad dependen de p. La varianza será máxima cuando:

$$\frac{d[p(1 - p)]}{dp} = 1 - 2p = 0$$

que implica $p = 0{,}5$. En este caso existe la mayor incertidumbre respecto al resultado y la mayor variabilidad: aparecerán a largo plazo el mismo número de ceros que de unos. Por el contrario, si p es muy pequeño (o muy grande), casi siempre obtendremos un uno (o un cero) y la variabilidad será menor.

5.1.3 Distribución binomial

La variable binomial se define en un proceso de Bernoulli por:

> y = número de elementos defectuosos al observar n

El *espacio muestral* de y o conjunto de valores posibles son los valores 0, 1, ..., n.

Para calcular la probabilidad de un valor particular r, consideremos el suceso r elementos defectuosos, seguidos de $n - r$ aceptables, que representaremos

$$\underbrace{DD \ldots D}_{r} \qquad \underbrace{A \ldots A}_{n-r}$$

Por la hipótesis de independencia, la probabilidad de este suceso es:

$$\underbrace{p \ldots p}_{r} \qquad \underbrace{(1-p) \ldots (1-p)}_{n-r} = p^r(1-p)^{n-r}$$

La probabilidad de r elementos defectuosos en cualquier orden requiere sumar las probabilidades de todos los sucesos excluyentes que verifican esta condición. Estos sucesos se obtienen permutando las letras anteriores de todas las formas posibles. Su número es igual a las permutaciones de n elementos con r y $n - r$ repetidos; este número es:

$$\frac{n!}{r!(n-r)!} = \binom{n}{r}$$

Por tanto:

$$P(y = r) = \binom{n}{r} p^r q^{n-r} \qquad ; \qquad r = 0, 1, \ldots, n \tag{5.4}$$

es fácil comprobar (véase el apéndice 5A) que:

$$E[y] = \Sigma rP(y = r) = np \tag{5.5}$$

$$DT[y] = \sqrt{\Sigma(r - np)^2 P(y = r)} = \sqrt{npq} \tag{5.6}$$

La tabla 2 del apéndice proporciona probabilidades binomiales acumuladas, y la figura 5.1 presenta cuatro ejemplos de la distribución. La asimetría de la distribución aumenta con la diferencia $q - p$ y la distribución es simétrica para $p = q = 0{,}5$.

5.1.4 Distribución geométrica

Consideremos el mismo mecanismo de generación de sucesos que en el modelo binomial, pero en lugar de contar el número de defectos en una muestra de n, consideremos:

x = número de elementos hasta el primer defectuoso

Para calcular su función de probabilidades, observemos que x tomará el valor n únicamente en el suceso:

$$\underset{n-1}{A \ldots A} \quad D$$

Por tanto, por la independencia

$$P(x = n) = p(1 - p)^{n-1};\ n = 1,\ 2,\ldots \tag{5.7}$$

Observemos que a diferencia de la variable binomial, el conjunto de valores posibles de la variable geométrica es ilimitado. Sin embargo:

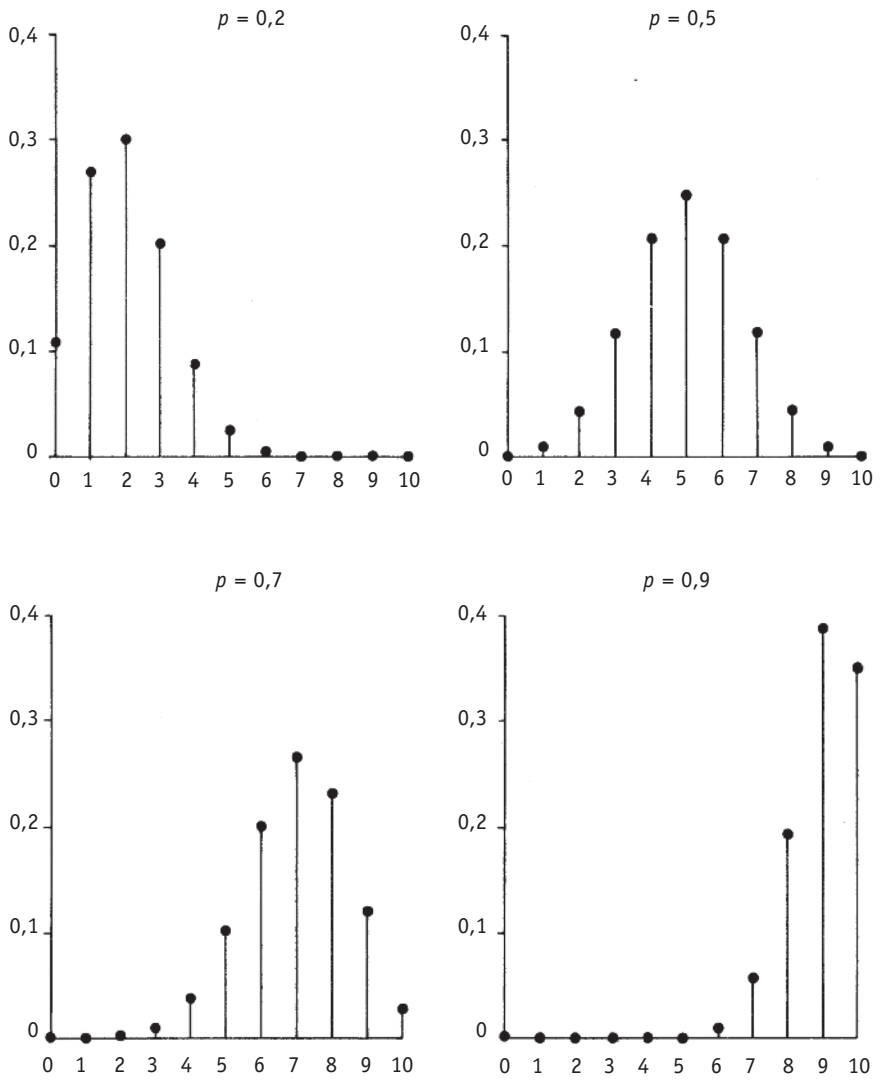
$$\sum_{1}^{\infty} P(x = n) = p \sum_{1}^{\infty} (1 - p)^{n-1} = 1$$

aplicando la fórmula de la suma de una progresión geométrica indefinida. La media y la desviación típica de esta distribución se calculan de la forma habitual (apéndice 5A) resultando:

$$E[x] = 1/p \tag{5.8}$$

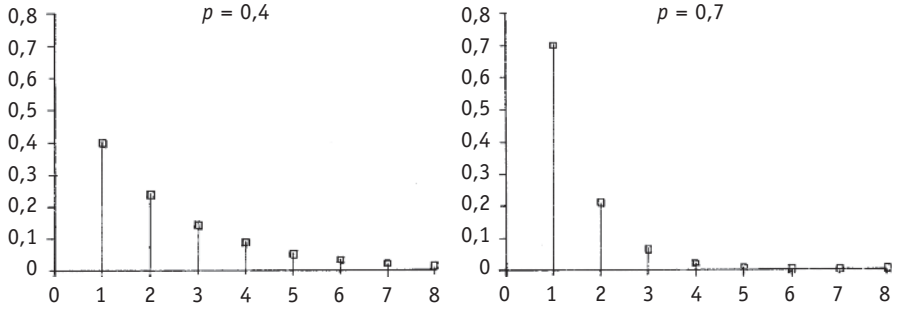
$$\text{Var}\,[x] = q/p^2 \tag{5.9}$$

Figura 5.1 Gráficos de barras para la distribución binominal con n = 10 y varios valores de p



Estos resultados son intuitivamente lógicos: como al tirar un dado el uno tiene probabilidad 1/6, se requerirán, por término medio, seis tiradas para que el uno aparezca. La fórmula (5.9) indica que la desviación promedio de este valor es menor que seis tiradas. La figura 5.2 presenta dos ejemplos de esta distribución.

Figura 5.2 Dos ejemplos de la distribución geométrica



Ejemplo 5.1

Supongamos un sistema con 9 componentes que requiere para su funcionamiento que al menos 6 estén disponibles. Si la probabilidad de funcionamiento de un componente es 0,95, calcular la fiabilidad del sistema (probabilidad de que funcione).

La probabilidad de que estén disponibles 6 o más componentes será:

$$P(\text{funcione}) = R = \sum_{6}^{9} \binom{9}{i} 0{,}95^{i} 0{,}05^{9-i} = \binom{9}{6} 0{,}95^{6} 0{,}05^{3} + \binom{9}{7} 0{,}95^{7} 0{,}05^{2} +$$

$$+ \binom{9}{8} 0{,}95^{8} \cdot 0{,}05 + \binom{9}{9} 0{,}95^{9} = 0{,}9917$$

Ejemplo 5.2

Un contrato de compra estipula la compra de componentes en lotes grandes que deben contener un máximo del 10% de componentes con algún defecto. Para comprobar la calidad se inspeccionan 11 unidades del producto con reposición de cada lote, aceptándolo si hay como máximo una unidad defectuosa. Estudiar cómo varía la probabilidad de aceptación por lote cuando la proporción real de componentes con algún defecto en los lotes es de 0,05, 0,10, 0,15 y 0,20. Conclusiones:

Si la proporción de defectos es p, la probabilidad de aceptar el lote será:

$$P(\text{aceptar}) = \sum_{i=0}^{1} \binom{11}{i} p^{i} q^{11-i}$$

suponiendo que la inspección se hace con reposición o que el lote es muy grande de manera que si no hay reposición no cambie.

Por ejemplo, para $p = 0{,}05$:

$$P(\text{aceptar}/p = 0{,}05) = \binom{11}{0} 0{,}05^0 \cdot 0{,}95^{11} + \binom{11}{1} 0{,}05 \cdot 0{,}95^{10} = 0{,}8981$$

Análogamente, para otros valores de p con ayuda de la tabla 2 del apéndice que proporciona las probabilidades binomiales se obtiene:

p	0,05	0,10	0,15	0,20	0,25
$P(\text{aceptar}/p)$	0,8981	0,06974	0,4922	0,3221	0,1971

Se observa que el plan establecido es desfavorable para el vendedor: se rechazarán el 30% de los lotes con calidad igual a la asegurada. Tampoco es muy adecuado para el comprador, que aceptará el 49% de las veces lotes con calidad peor que la establecida ($p = 0{,}15$). Para mejorar el plan de control habrá que aumentar n, número de piezas inspeccionadas, como el lector debe comprobar.

Ejemplo 5.3

Una pareja decide tener hijos hasta el nacimiento de la primera niña. Calcular la probabilidad de que tengan más de 4 hijos. [Tomar $P(\text{niño}) = P(\text{niña}) = 0{,}5$].

Sea X = número de hijos antes de la primera niña. Entonces X es una variable geométrica; por tanto:

$$P(X > 4) = 1 - P(X \leqslant 4) = 1 - \sum_{i=1}^{4} (1/2)^i = 0{,}0625$$

5.2 El proceso de Poisson y sus distribuciones asociadas

5.2.1 El proceso de Poisson

Consideremos un experimento en el que observamos la aparición de sucesos puntuales sobre un soporte continuo. Por ejemplo, averías de máquinas en el tiempo, llegadas de aviones a un aeropuerto, pedidos a una empresa,

estrellas en el firmamento en cuadrículas del mismo tamaño, defectos en una plancha de metal, etc. Supondremos que el proceso que genera estos sucesos se caracteriza por:

1) Es estable: produce, a largo plazo, un número medio de sucesos constante por unidad de observación (tiempo, espacio, área, etc.).
2) Los sucesos aparecen aleatoriamente de forma independiente, es decir, el proceso no tiene memoria: conocer el número de sucesos en un intervalo no ayuda a predecir el número de sucesos en el siguiente.

Este proceso es la generalización a un soporte continuo del proceso de Bernoulli.

5.2.2 La distribución de Poisson

La variable de Poisson se define en el proceso anterior como:

x = número de sucesos en un intervalo de longitud fija

La distribución de Poisson aparece como límite de la distribución binomial si suponemos que el número de elementos observados es muy grande pero que la probabilidad de observar la característica estudiada en cada elemento es muy pequeña. En efecto, dividamos el intervalo de observación, *t,* en *n* segmentos muy pequeños (de manera que *n* será muy grande) y observemos en cada segmento si ocurre o no el suceso estudiado. Si la probabilidad de este suceso en cada segmento, *p*, es muy pequeña, la aparición de dos o más sucesos en un segmento será despreciable, y podemos plantear el problema como observar en *n* elementos (segmentos) si aparece o no el suceso estudiado. Ésta es la distribución binomial, y es claro que la distribución de Poisson corresponderá a un caso límite de ésta cuando *n* tienda a infinito y *p* tienda a cero, pero de manera que el número medio esperado de sucesos, *np,* permanezca constante.

Consideremos, por ejemplo, que el número de accidentes por 100 horas de conducción para un grupo de conductores es λ, y que los accidentes ocurren de acuerdo con el proceso de Poisson: aleatoria e independientemente a lo largo del tiempo. La variable *x*, número de accidentes en 100 horas de conducción, será una variable de Poisson. Para obtener su distribución, observemos que podemos convertir *x* en binomial considerando intervalos de tiempo muy pequeños (por ejemplo, cada minuto), donde la probabilidad de dos accidentes sea despreciable. Entonces, *x* puede considerarse como una variable binomial en un experimento con

$n = 100 \cdot 60 = 6.000$ repeticiones, cada una consistente en observar si en un minuto ha ocurrido o no un accidente. La probabilidad de accidente p será tal que:

$$E(x) = \lambda = np \quad ; \quad p = \lambda/n$$

Por tanto, disminuyendo el intervalo de observación (aumentando n) pero manteniendo $\lambda = np$ se obtendrá la distribución de Poisson.

Análogamente, la variable binomial número de piezas defectuosas en una cadena que produce un gran número de ellas puede aproximarse, cuando n sea muy grande y p muy pequeña, por la variable que cuenta el número de defectos por intervalo de tiempo, que es la variable de Poisson con $\lambda = np$.

En conclusión, las probabilidades de Poisson pueden aproximarse por

$$P(x = r) = \binom{n}{r} \left(\frac{\lambda}{n}\right)^r \left(1 - \frac{\lambda}{n}\right)^{n-r},$$

y tomando límites:

$$\lim_{n\to\infty} P(x = r) = \frac{\lambda^r}{r!} \lim_{n\to\infty} \frac{n(n-1) \ldots (n-r+1)}{\left(1 - \frac{\lambda}{n}\right)^r n^r} \left(1 - \frac{\lambda}{n}\right)^n$$

$$\lim_{n\to\infty} \frac{n}{(n-\lambda)} \cdot \frac{(n-1)}{(n-\lambda)} \ldots \frac{(n-r+1)}{(n-\lambda)} = 1$$

$$\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$$

Tendremos que, en el límite

$$P(x = r) = \frac{\lambda^r}{r!} e^{-\lambda} \qquad r = 0, 1, 2, \ldots \tag{5.10}$$

que es la distribución de Poisson. Sus medidas características son (véase apéndice 5A).

$$E[x] = \sum_0^\infty r \frac{\lambda^r}{r!} e^{-\lambda} = \lambda e^{-\lambda} \sum_1^\infty \frac{\lambda^{r-1}}{(r-1)!} = \lambda \tag{5.11}$$

$$\text{Var}[x] = \Sigma(r - \lambda)^2 \frac{\lambda^r}{r!} e^{-\lambda} = \lambda \tag{5.12}$$

Observemos que estos resultados son consistentes con la aproximación binomial: la varianza de la binomial es *npq* y cuando $n \to \infty$ y $p \to 0$, pero $np = \lambda = cte$, entonces $q \to 1$, lo que implica

$$npq \to np = \lambda$$

que es la varianza de Poisson.

Las probabilidades acumuladas de la distribución de Poisson para distintos valores de λ están en la tabla 3 del apéndice tablas. La distribución es asimétrica, pero tiende a la simetría al aumentar λ. La figura 5.3 presenta ejemplos de esta distribución.

Ejemplo 5.4

Las llamadas por averías en un puesto de servicio siguen una distribución de Poisson de media dos averías/semana. Calcular la probabilidad de:

a) Ninguna avería en una semana.
b) Menos de cinco en una semana.
c) Menos de seis en un mes (cuatro semanas).

a) $p(x = 0) = \dfrac{2^0}{0!} \cdot e^{-2} = 0{,}14$

b) Utilizando la tabla 3 del apéndice se obtiene:

$$p(x \leq 4) = e^{-2}\left(1 + 2 + \frac{2^2}{2!} + \frac{2^3}{3!} + \frac{2^4}{4!}\right) = 0{,}947$$

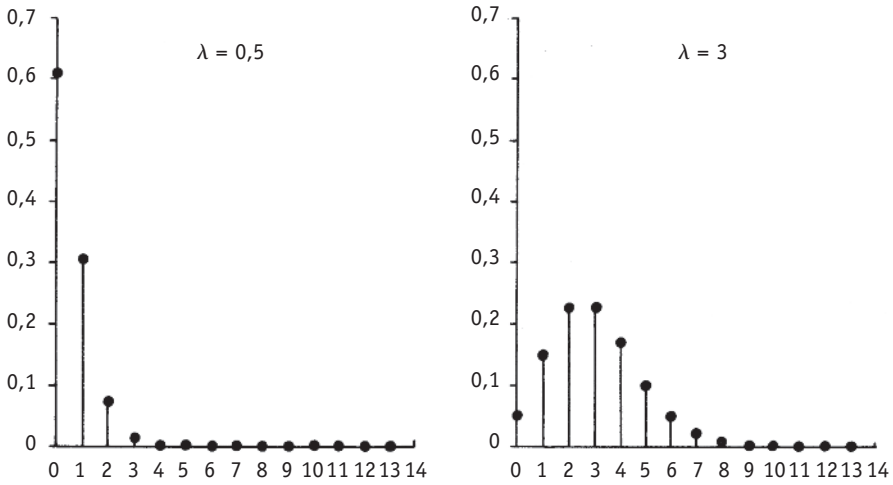
c) La media de cuatro semanas será $\lambda' = 4 \cdot \lambda = 8$ averías/4 semanas. Entonces:

$$p(x \leq 5/\lambda' = 8) = e^{-8}\left(1 + 8 + \frac{8^2}{2!} + \frac{8^3}{3!} + \frac{8^4}{4!} + \frac{8^5}{5!}\right) = 0{,}191$$

5.2.3 Distribución exponencial

La variable exponencial resulta al considerar en un proceso de Poisson la variable continua

Figura 5.3 Gráficos de barras para la distribución de Poisson con dos valores de λ



t = tiempo entre la ocurrencia de dos sucesos consecutivos

que tomará valores en el intervalo $(0, \infty)$. Para obtener su función de distribución, observemos que

$$P(t > t_0) = P[\text{cero sucesos en intervalo } (0, t_0)] = e^{-\lambda t0},$$

siendo λ la tasa media de sucesos por unidad de tiempo. Entonces:

$$F(t_0) = P(t \leq t_0) = 1 - e^{-\lambda t0}$$

cuya función de densidad será:

$$f(t) = \frac{dF(t)}{dt} = \lambda e^{-\lambda t} \quad ; \quad \lambda > 0, t > 0 \tag{5.13}$$

Las medidas características de esta distribución son:

$$E[t] = \frac{1}{\lambda} = DT[t] \tag{5.14}$$

Por ejemplo, si la tasa de llegadas de clientes a un puesto de servicio es de cuatro clientes/hora, el tiempo medio entre clientes es 1/4 de hora o quince minutos, y éste resulta ser el valor de la desviación típica.

Esta distribución es el equivalente continuo de la geométrica: si el tiempo medio entre clientes es 15 minutos y observamos cada minuto si llega o no un cliente, la probabilidad de llegada en cada minuto es 1/15. La distribución geométrica estudia el número de observaciones (minutos) promedio entre llegadas de clientes, que será 15, igual al valor promedio para la exponencial.

La similitud entre las fórmulas (5.8) y (5.14) es consecuencia de esta analogía. La figura 5.4 presenta esta distribución. La desviación típica es en general mayor en la distribución exponencial (identificando λ en el caso continuo con p en el discreto), y la diferencia disminuye con p. Ambas distribuciones decrecen muy lentamente si p es muy pequeño (λ pequeño).

Ejemplo 5.5

Se ha comprobado que la duración de vida de ciertos elementos sigue una distribución exponencial con media 8 meses. Se pide: 1) calcular la probabilidad de que un elemento tenga una vida entre 3 y 12 meses; 2) el percentil 0,95 de la distribución; 3) la probabilidad de que un elemento que ha vivido ya más de 10 meses viva más de 25.

1) Como la media es $1/\lambda$, tendremos que

$$f(t)\ \frac{1}{8}\ e^{-t/8} \qquad t > 0$$

donde t va medido en meses. Entonces:

$$P(3 < t < 12) = \int_3^{12} \frac{1}{8}\, e^{-t/8} dt = -\, e^{-t/8} \Big]_3^{12} = 0{,}69 - 0{,}22 = 0{,}47$$

2) $$F(x) = \int_0^x f(t)dt = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}$$

$$0{,}95 = 1 - e^{-x/8}$$

$$x = -\,8 \ln 0{,}05 = 23{,}97$$

3) La probabilidad pedida es

$$P(t > 25 \mid t > 10)$$

y por definición de probabilidad condicionada:

$$P(t > 25 \mid t > 10) = \frac{P(t > 25)}{P(t > 10)} = \frac{1 - P(t \leqslant 25)}{1 - P(t \leqslant 10)}$$

ya que la probabilidad conjunta $P(t > 25,\ t > 10)$ se reduce a $P(t > 25)$ al estar el primer suceso contenido en el segundo. Por tanto, utilizando la expresión anterior de la función de distribución:

$$P(t > 25 \mid t > 10) = \frac{1 - (1 - e^{-\lambda 25})}{1 - (1 - e^{-\lambda 10})} = e^{-\lambda(25-10)} = e^{-\lambda 15}$$

Por tanto:

$$P(t > 25 \mid t > 10) = P(t > 15)$$

y en general, es inmediato ver que si $t_2 > t_1$

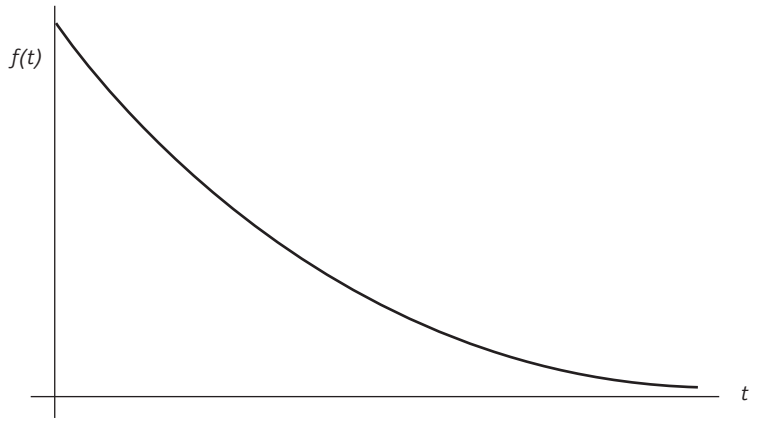
$$P(t > t_2 \mid t > t_1) = P(t > t_2 - t_1)$$

que indica que la probabilidad de que un elemento viva $t_2 - t_1$ unidades de tiempo adicionales es independiente del tiempo ya vivido por el elemento. En este sentido se dice que la distribución exponencial no tiene memoria.

5.3 Distribuciones de duraciones de vida

La distribución exponencial es el ejemplo más simple de las distribuciones para variables aleatorias continuas que pueden tomar cualquier valor positivo no acotado. Estas distribuciones se utilizan para modelar la duración (vida de personas, animales o componentes físicos; duración de huelgas, período de desempleo, etc.) o el tamaño (rentas de familias, duración de discursos políticos, tamaño de yacimientos, etc.). Para concretar, supondremos que la variable de interés es la duración de vida de ciertos elementos. Una forma de caracterizar estas distribuciones es por la función que proporciona la probabilidad de muerte en cada instante para los elementos que han sobrevivido hasta dicho instante. Esta función se denomina *tasa de fallo,* y define la función de densidad de la variable.

Figura 5.4 La distribución exponencial



En efecto, sea $f(t)$ la función de densidad de una variable continua positiva en $(0, \infty)$, la probabilidad de muerte en el intervalo $(t_0, t_0, + \Delta t)$ para los elementos que ya han vivido t_0 es, aplicando la definición de probabilidad condicionada:

$$P(t_0 < t \leq t_0 + \Delta t | t > t_0) = \frac{P(t_0 < t \leq t_0 + \Delta t)}{P(t > t_0)}$$

ya que la probabilidad conjunta de los sucesos $t > t_0$ y $t_0 < t \leq t_0 + \Delta t$ coincide con la probabilidad del segundo. Llamando $F(t_0)$ a la función de distribución de la variable en t_0:

$$P(t_0 < t \leq t_0 + \Delta t | t > t_0) = \frac{f(t_0)\Delta t}{1 - F(t_0)}$$

y en el límite, se define la tasa de fallo, $\lambda(t)$, por:

$$\boxed{\lambda(t) = \frac{f(t)}{1 - F(t)}} \qquad (5.15)$$

Para obtener la función de densidad en función de la tasa de fallo, integrando (5.15) entre 0 y t; y teniendo en cuenta que $F(0) = 0$:

$$\int_0^t \lambda(x)\, dx = \int_0^t \frac{f(x)}{1 - F(x)}\, dx = -\ln[1 - F(x)]_0^t = -\ln[1 - F(t)]$$

y llamando:

$$\wedge(t) = \int_0^t \lambda(x)\, dx \tag{5.16}$$

a la función de tasas de fallo acumulada, obtenemos que:

$$1 - F(t) = \exp\{-\wedge(t)\}$$
$$F(t) = 1 - \exp\{-\wedge(t)\}$$

de donde resulta, derivando la función de densidad

$$\boxed{f(t) = \lambda(t) \exp\{-\wedge(t)\}} \tag{5.17}$$

que es la forma habitual de las distribuciones continuas para variables positivas.

La distribución exponencial se caracteriza por una tasa de fallo constante: la probabilidad de morir en cualquier intervalo no depende de la vida anterior. Resulta, por tanto, adecuada para describir la aparición de muertes al azar, no debidas a desgaste o deterioro.

Si suponemos que la tasa de fallo es del tipo:

$$\lambda(t) = ht^{c-1}$$

tendremos que la tasa de fallo aumentará con el tiempo si $c > 1$, será constante (distribución exponencial) si $c = 1$ y disminuirá si $c < 1$. Entonces la función de densidad será:

$$f(x) = ht^{c-1} \exp\left\{-\frac{h}{c} t^c\right\}$$

que se conoce como *distribución de Weibull*.

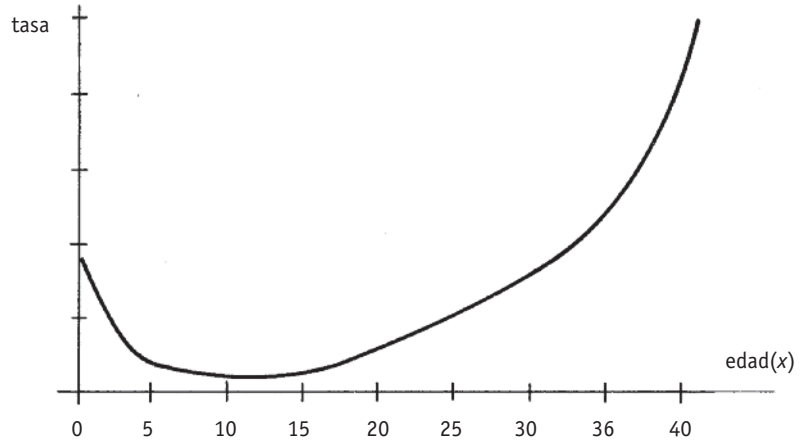
La tabla 5.1 recoge las probabilidades de muerte y tasas de fallo (mortalidad) para la población española en 1986. Suponiendo estabilidad, las probabilidades anuales de muerte por grupos de edad se han obtenido dividiendo el número de defunciones en un año en cada grupo de edades por la población en dicho grupo de edad. La tasa de mortalidad o tasa de fallo anual de un grupo de edades se obtiene dividiendo las defunciones de ese grupo de edades en un año por el número de personas con edad superior o igual al grupo considerado.

La figura 5.5 representa la evolución de la tasa de mortalidad de la población española. Esta curva en forma de bañera es característica de los estudios de duraciones de vida y refleja tres tramos bien diferenciados:

Tabla 5.1 Población (censo 1981) y defunciones en 1983 en España (datos INE)

Clase	Censo 1981 Población en miles	Defunciones en 1983	1.000 × Prob. de muerte	1.000 × Tasa mortalidad
Menores 5 años	3.075	6.584	2,14	0,175
5 - 9	3.308	896	0,27	0,026
10 - 14	3.302	869	0,26	0,028
15 - 24	6.205	4.161	0,67	0,149
25 - 34	4.993	4.432	0,89	0,203
35 - 44	4.302	6.860	1,59	0,408
45 - 54	4.626	18.107	3,91	1,449
55 - 64	3.634	37.081	10,20	4,710
65 o más	4.237	223.579	52,76	52,755
TOTAL	37.683	302.569		

Figura 5.5 Tasa de mortalidad (fallo) de la población española a partir de los datos de la tabla



a) un primer tramo de tasa de mortalidad decreciente, que es debida a la alta mortalidad relativa en el parto;
b) un tramo de mortalidad constante, hasta la adolescencia;
c) un crecimiento exponencial de la mortalidad desde entonces.

El primer tramo puede representarse por una Weibull con $c < 1$, el segundo por una exponencial y el tercero por una distribución con tasa de fallo:

$$\lambda(t) = ke^{bt}$$

que crezca exponencialmente con el tiempo, con lo que la distribución resultante para ese tercer tramo es:

$$f(x) = ke^{bt} \exp\left\{-\frac{k}{b}e^{bt} + \frac{k}{b}\right\} \tag{5.18}$$

que se conoce como *distribución de Gompertz,* y se utiliza mucho en estadística actuarial porque proporciona una descripción bastante precisa de la duración de la vida humana después de los 20 años.

5.4 La distribución normal

El modelo de distribución de probabilidad para variables continuas más importante es la distribución normal, cuya función de densidad es

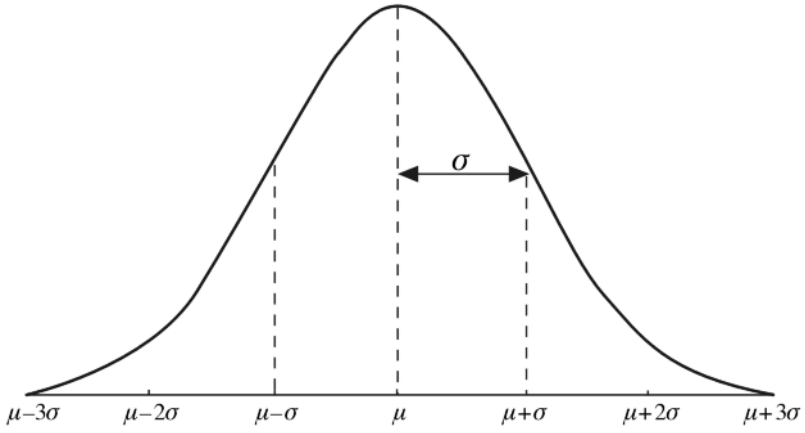
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right] \tag{5.19}$$

que aparece dibujada en la figura 5.6. La función f depende de dos parámetros: μ, que es al mismo tiempo la media, la mediana y la moda de la distribución, y σ, que es la desviación típica. Diremos que una variable es $N(\mu, \sigma)$ cuando sigue la función de densidad (5.19).

La distribución normal aproxima lo observado en muchos procesos de medición sin errores sistemáticos. Por ejemplo, las medidas físicas del cuerpo humano en una población, las características psíquicas medidas por test de inteligencia o personalidad, las medidas de calidad en muchos procesos industriales o los errores de las observaciones astronómicas siguen distribuciones normales. Una justificación de la frecuente aparición de la distribución normal es el *teorema central del límite* que veremos en la sección siguiente y que establece que cuando los resultados de un experimento son debidos a un conjunto muy grande de causas independientes, que actúan sumando sus efectos, siendo cada efecto individual de poca importancia respecto al conjunto, es esperable que los resultados sigan una distribución normal.

La variable normal con $\mu = 0$ y $\sigma = 1$ se denomina normal estándar, $N(0, 1)$, y su función de distribución está tabulada (véase la tabla 4 del apéndi-

Figura 5.6 Distribución normal



ce). Para calcular probabilidades en el caso general, transformaremos la variable aleatoria normal x en la variable normal estándar z, mediante:

$$z = \frac{x - \mu}{\sigma}$$

que convierte una variable x con media μ y desviación típica σ en la normal estándar z. En efecto, utilizando la fórmula (4.34) para el cambio de variable:

$$f(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad \sigma = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

que es la normal estándar. El cálculo de probabilidades de x se efectúa utilizando la expresión:

$$F(x_0) = P(x \le x_0) = P(\mu + \sigma z \le x_0) = P\left(z \le \frac{x_0 - \mu}{\sigma}\right) = \Phi\left(\frac{x_0 - \mu}{\sigma}\right)$$

donde $\Phi(\cdot)$ representa la función de distribución de la normal estándar (véase la tabla 4 en el apéndice de tablas). Esta expresión indica que podemos calcular el valor de la función de distribución de cualquier variable normal en cualquier punto si conocemos la función de distribución de la normal estándar. Sólo tenemos que convertir el punto x_0 en un punto de la normal estándar restándole la media y diviendo por la desviación típica.

Se comprueba que, en toda distribución normal, en el intervalo:

$\mu \pm 2\sigma$ se encuentra el 95,5% de la distribución
$\mu \pm 3\sigma$ encuentra el 99,7% de la distribución

Conocer que unos datos siguen una distribución normal nos permite dar intervalos más precisos que los de la acotación de Tchebychev.

La distribución normal se toma como referencia para juzgar muchas otras distribuciones. Por ejemplo, el coeficiente de apuntamiento de la normal es 3, y algunos programas de ordenador calculan el coeficiente de apuntamiento de cualquier distribución como:

$$CA_p = \frac{\mu_4}{\sigma^4} - 3$$

de manera que para la normal sea cero y el signo indique un mayor o menor apuntamiento respecto a ésta. Los cuartiles de una distribución normal son (véase tabla 4) $-0{,}675\sigma$, 0, $0{,}675\sigma$, lo que implica que el rango intercuartílico es $1{,}35\sigma$ y la Meda $0{,}675\sigma$.

Ejemplo 5.6

Una de las primeras aplicaciones de la curva normal fue debida al astrónomo F. W. Bessel en 1818, que comprobó que los errores de medida de 300 medidas astronómicas coincidían con bastante aproximación con los previstos por Gauss con la curva normal. Suponiendo que la media de estos errores es cero y la desviación típica 4 grados, calcular: 1) la probabilidad de que un error no sea mayor que 6 grados; 2) la probabilidad de que sea por defecto y mayor que 8 grados; 3) si llamamos «pequeños» a los errores menores que 7 grados y «grandes» a los mayores que 7 grados, calcular el número esperado de errores grandes y pequeños en 300 observaciones.

1) Sea x la variable normal que refleja la distribución de los errores, $x \sim N(0,4)$ y z la $N(0,1)$. Entonces:

$$P(|x| \leqslant 6) = P(-6 \leqslant x \leqslant 6) = P(-6/4 \leqslant z \leqslant 6/4) = \phi(1{,}5) - \phi(-1{,}5) =$$
$$= 0{,}93319 - 0{,}06681 = 0{,}86638.$$

2) $P(x < -8) = P(x/4 < -2) = P(z < -2) = 0{,}02275.$

3) Sea *a* = pequeño; *B* = grande. Entonces:

$$P(A) = P(|x| \leqslant 7) = P(|z| \leqslant 1{,}75) = 0{,}95994 - 0{,}04006 = 0{,}91988.$$

$$P(B) = 1 - 0{,}91988 = 0{,}080012.$$

Para calcular el número esperado de observaciones *A* en 300, observemos que cada error puede ser grande o pequeño con probabilidad constante. Suponiendo que los errores son independientes unos de otros, tendremos una distribución binomial con $n = 300$ y $P(A)$ constante. Por tanto:

$$E[\text{n.º de casos } A] = 300\ P(A) = 275{,}96 \simeq 276$$

$$E[\text{n.º de casos } B] = 300\ P(B) = 24{,}04 \simeq 24$$

5.5 La normal como aproximación de otras distribuciones

5.5.1 El teorema central del límite

Este teorema establece que si $x_1, ..., x_n$ son variables aleatorias independientes con media μ_i y varianza σ_i^2 y distribución cualquiera —no necesariamente la misma— y formamos la variable suma

$$Y = x_1 + ... + x_n \tag{5.20}$$

entonces, si cuando *n* crece $\sigma_i^2/\Sigma\sigma_j^2 \to 0$, que implica que el efecto de una variable es pequeño respecto al efecto total, la variable

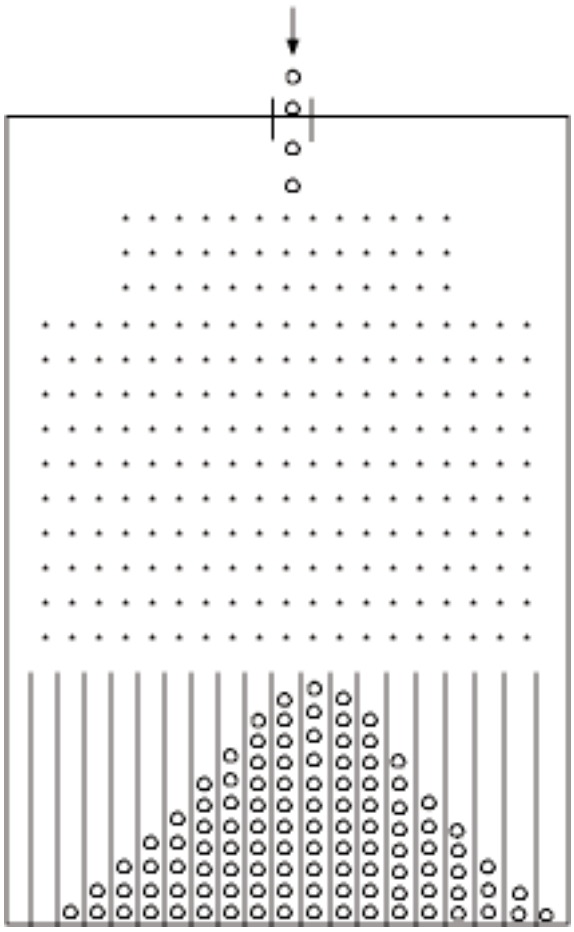
$$\frac{Y - \Sigma\mu_i}{\sqrt{\Sigma\sigma_i^2}}$$

tiende hacia una distribución $N(0,1)$. El resultado anterior implica que si *n* es grande, podemos aproximar las probabilidades de *Y* utilizando que:

$$Y \sim N\left(\sum \mu_i; \sqrt{\sum \sigma_i^2}\right) \tag{5.21}$$

En este teorema aparecen tres resultados distintos. El primero, que si tenemos una variable que es suma de otras, la media de la suma es la suma de las medias. El segundo, que cuando los sumandos son independientes, la varianza de la variable suma es la suma de las varianzas de los sumandos. Estos dos resultados son siempre ciertos y los demostraremos en el capítulo siguiente. El tercer resultado es la clave del teorema: la variable suma se distribuye normalmente. De acuerdo con este teorema, siempre que observemos una varia-

Figura 5.7 Aparato construido por Galton para comprobar el teorema central del límite



ble que sea el resultado de muchas causas independientes que se suman, esperamos que su distribución sea aproximadamente normal. Por ejemplo, las medidas físicas (altura, longitud de los brazos o piernas, etc.) de una persona son debidas a muchas causas distintas (herencia genética pero también alimentación, ejercicio, hábitos infantiles, etc.) y, esperamos que sigan una distribución normal. Lo mismo ocurre con el grado de acuerdo de una población (en una escala de 0 a 100 por ejemplo) con temas no conflictivos y sin gran carga emocional, o con las fluctuaciones de la demanda por un producto o servicio cuando la demanda es estable en el tiempo y las fluctuaciones entre períodos se deben a la suma de muchas causas pequeñas.

Galton tuvo la intuición de construir el aparato que se presenta esquemáticamente en la figura 5.7. Por la parte superior se introducen bolitas que des-

cienden chocando con los palitos que cubren todo su recorrido y en el fondo se van depositando en pequeños carriles. Se comprueba que la distribución de las bolitas en los carriles reproduce aproximadamente la distribución normal. Esto es esperable, ya que la desviación que sufren en su trayectoria depende de muchas pequeñas causas (los choques) que actúan sumándose. Veremos en la sección 5.7 otros métodos para comprobar el teorema central del límite simulando la generación de variables con un ordenador.

Ejemplo 5.7

Se dispone de 100 números con cuatro decimales. Para obtener su suma, los números se convierten en enteros por redondeo. Calcular la probabilidad de que el error de redondeo cometido sea mayor que cinco unidades.

Solución:

Sean o_i los números originales y r_i los redondeados. Se verifica:

$$r_i = o_i + u_i$$

donde u_i es una variable de media cero que toma valores en (–0,5; 0,5) y que supondremos sigue una distribución uniforme. Entonces:

$$e = \Sigma r_i - \Sigma o_i = \Sigma u_i$$

Como el valor medio de u_i es cero y su varianza es (cuadro 3.2) 1/12, tendremos que la variable e es aproximadamente normal con media cero y varianza $100/12 \simeq 8{,}3$. Por tanto:

$$P(|e|>5) = 1-P(|e|\leq 5)=1-P\left(\left|\frac{e}{\sqrt{8{,}3}}\right| \leq \frac{5}{\sqrt{8{,}3}}\right) = 1-P(|z|\leq 1{,}73)=0{,}0836$$

donde z es una variable aleatoria normal estándar.

5.5.2 Relación entre binomial, Poisson y normal

La variable binomial Y definida en la sección 5.1 es la suma de n variables de Bernoulli, x_i, que toman el valor 1 cuando el elemento es defectuoso y cero en caso contrario. Entonces:

$$Y = x_1 + ... + x_n$$

donde $x_i = 1$ si el *i*-ésimo elemento es defectuoso. Estamos pues en un caso particular del teorema central del límite. Como $E[x_i] = p$, $\text{Var}[x_i] = pq$, la variable Y tenderá hacia la normal con parámetros np y $\sqrt{npq}$.

En 1733, De Moivre demostró este resultado buscando cómo aproximar las probabilidades binomiales. Este autor encontró que si x es una variable binomial de parámetro p, la distribución de

$$\frac{x - np}{\sqrt{npq}}$$

converge hacia una distribución normal con media cero y varianza uno. En la práctica, esto se traduce en que si n es grande (mayor que 30), y p no muy cercano a cero o uno, podemos calcular la probabilidad de que la variable binomial x esté en (a, b) considerando a x como una variable normal, de $\mu = np$ y $\sigma = \sqrt{npq}$, y buscando el área encerrada entre a y b. La aproximación mejora tomando el intervalo $(a - 0{,}5;\ b + 0{,}5)$, que tiene en cuenta que el número entero n equivale al intervalo continuo $(n - 0{,}5;\ n + 0{,}5)$. Por tanto, la condición para una variable discreta

$$a \leq x \leq b$$

equivale, para una variable continua, a:

$$a - 0{,}5 \leq x \leq b + 0{,}5$$

En general esta aproximación se utiliza para $npq > 5$. La figura 5.8 ilustra gráficamente este resultado.

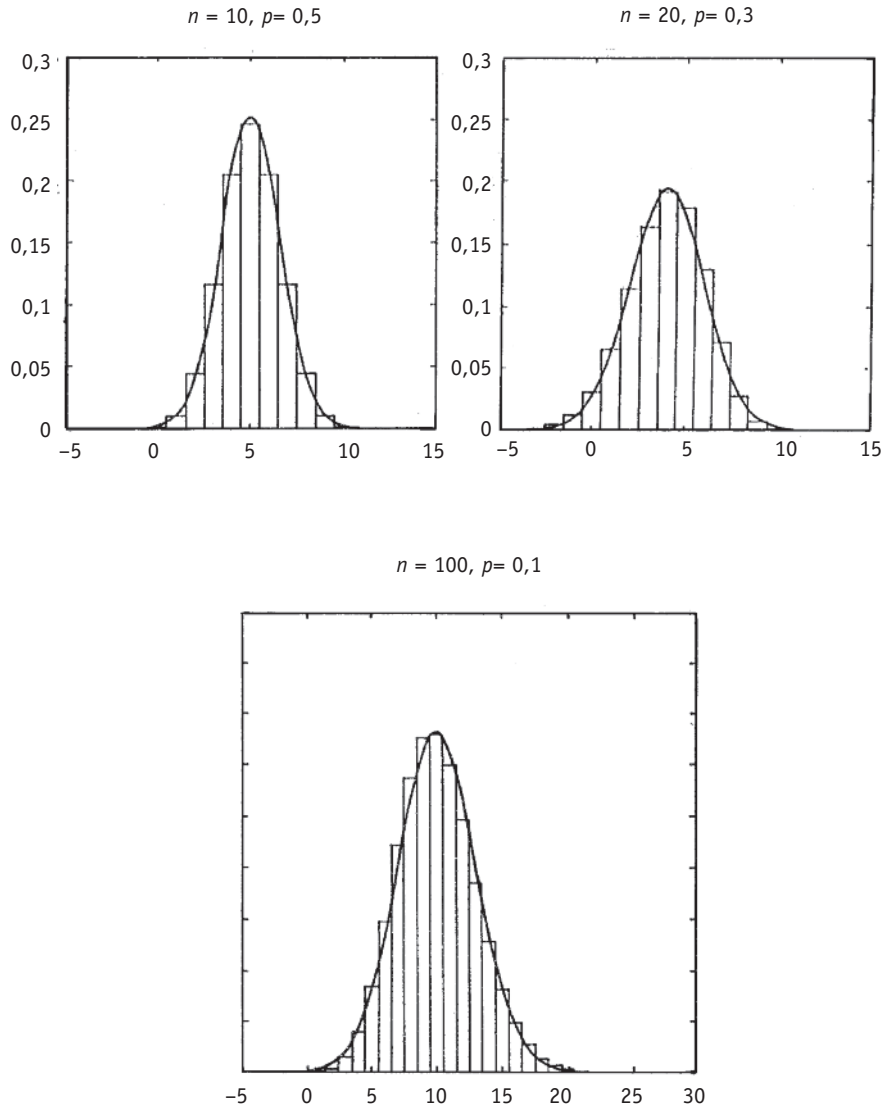
Esta misma situación aparece con variables de Poisson: sea $Y(0, T)$ la variable de Poisson que cuenta el número de sucesos en $(0, T)$. Dividiendo el intervalo en n partes iguales, esta variable puede expresarse como:

$$Y(0, T) = x_1(0, t_1) + x_2(t_1, t_2) + ... + x_n(t_{n-1}, T)$$

donde $x_i(t_{i-1}, t_i)$ cuenta el número de sucesos en el intervalo (t_{i-1}, t_i). Se verifican por tanto las condiciones del teorema central y cuando n aumenta —lo que requiere que λ sea grande— , la distribución de Poisson se aproximará por la normal.

La aproximación es buena cuando $\lambda > 5$. El procedimiento operativo es, como en el caso anterior, utilizar la corrección de continuidad y escribir:

Figura 5.8 Convergencia de la distribución binomial hacia la normal

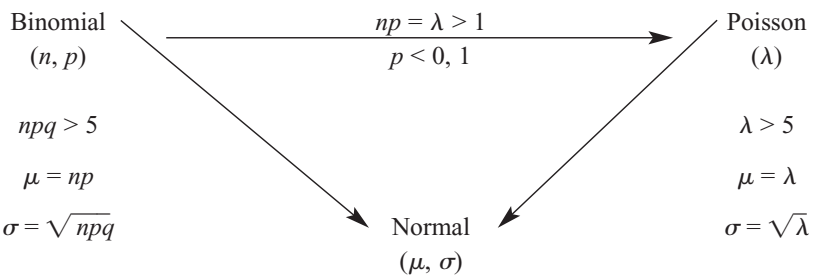


$$P(a \leq x_p \leq b) \simeq P(a - 0{,}5 \leq x_n \leq b + 0{,}5)$$

donde x_p es una variable de Poisson de parámetro λ y x_n es una variable normal de parámetros $\mu = \lambda$, $\sigma = \sqrt{\lambda}$.

El cuadro 5.1 resume estas aproximaciones.

Cuadro 5.1 Relación entre distribuciones



Ejemplo 5.8

En un proceso de fabricación de película fotográfica aparece por término medio un defecto por cada 20 metros de película. Si la distribución de defectos es Poisson, calcular la probabilidad de seis defectos en un rollo de 200 metros de película (a) directamente; (b) utilizando la aproximación normal.

Como $\lambda = \frac{1}{20}$ metro, en 200 metros, $\lambda = \frac{200}{20} = 10$ defectos/200 m.

$$P(x = 6) = \frac{e^{-10} \cdot 10^6}{6!} = 0{,}0630$$

con la normal

$$P(x=6)=P(5{,}5<x<6{,}5)=P\left(\frac{5{,}5-10}{\sqrt{10}} < z < \frac{6{,}5-10}{\sqrt{10}}\right)=0{,}9222-0{,}8665=0{,}0557$$

5.6 La distribución lognormal

Una consecuencia del teorema central del límite es que si un efecto es el *producto* de muchas causas cada una de poca importancia respecto a las demás e independientes, de manera que

$$y = x_1 x_2 \ldots x_n$$

entonces el logaritmo de y seguirá una distribución normal.

Se denomina distribución lognormal a la de una variable cuyo logaritmo se distribuye normalmente. Aplicando la fórmula (4.34), si:

$$x = \ln y$$

es normal $N(\mu, \sigma)$, la densidad de y será:

$$g(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\ln y - \mu}{\sigma}\right)^2\right\}\frac{1}{y} \qquad y > 0$$

Como la transformación logarítmica es monótona, los percentiles de x serán los logaritmos de los percentiles de y. Por ejemplo, la mediana de y será e^{μ}, siendo μ la media de la variable x. Los parámetros de la distribución lognormal se indican en el cuadro 3.2.

Esta distribución aparece con frecuencia al estudiar el tamaño de elementos: rentas de familias, consumo de electricidad por empresas, ventas en euros, etc.

La distribución lognormal es especialmente útil para comparar distribuciones asimétricas con variabilidad muy distinta. Es fácil demostrar, utilizando las fórmulas aproximadas (4.37) y (4.38) o las fórmulas exactas del cuadro 5.2, que si disponemos de varias poblaciones con distribución lognormal y el mismo coeficiente de variación (lo que equivale a decir que la desviación típica es siempre proporcional a la media), al tomar logaritmos obtenemos distribuciones aproximadamente asimétricas con la misma varianza. Esta varianza es el coeficiente de variación común de las distribuciones originales. Tomar logaritmos en estos casos simplifica mucho las comparaciones, ya que entonces las nuevas distribuciones sólo diferirán en sus medias.

Ejercicios 5.1

5.1.1. Un dado se lanza diez veces. Sea A el suceso un solo 6 en las diez tiradas; y B, exactamente dos veces 6. ¿Cuál de los dos es más probable?

5.1.2. ¿Qué es más probable, obtener un 6 al lanzar un dado una vez u obtener 3 seises al lanzar un dado seis veces?

5.1.3. Sea x una variable binomial (n, p). Sea $p_r = p(x = r)$. Demostrar que

$$p_r = \frac{n - r + 1}{r} \cdot \frac{p}{1 - p} P_{r-1}$$

y utilizar esta relación para generar las probabilidades $p_0, ..., p_6$ para una binomial con $n = 6$, $p = 1/3$.

5.1.4. Si las llamadas telefónicas a una centralita siguen una distribución de Poisson de parámetro $\lambda = 3$ llamadas/cinco minutos, calcular la probabilidad de:
a) Seis llamadas en cinco minutos.
b) Tres en diez minutos.
c) Más de 15 en un cuarto de hora.
d) Dos en un minuto.

5.1.5. Calcular en el problema anterior la probabilidad de que transcurran cinco minutos sin ninguna llamada.

5.1.6. En un libro de 200.000 palabras la probabilidad de que una palabra esté escrita incorrectamente es 1/50.000. Calcular:
a) La probabilidad de que no haya errores.
b) La probabilidad de más de seis errores.

5.1.7. Supongamos un experimento que tiene probabilidad de éxito igual a 0,01. Calcular cuántas veces debe repetirse para que la probabilidad de al menos tres éxitos sea como mínimo 0,9.

5.1.8. Supongamos un calculador que contiene cuatro circuitos impresos. Sea p_i la probabilidad de que un calculador enviado a reparar necesite i circuitos nuevos. Se conoce que $p_1 = 1/2$; $p_2 = 1/4$; $p_3 = p_4 = 1/8$. Se envían 10.000 unidades a reparar al año. ¿Cuál es la probabilidad de necesitar más de 18.875 circuitos?

5.1.9. En una marca de chocolates se incluyen cupones del 1 al 6. Determinar el número medio de paquetes necesarios para tener uno de cada tipo.

5.1.10. Una empresa recibe piezas de un proveedor en lotes de 2.000 que se someten al siguiente control de calidad: se toman 20 al azar y si hay más de una defectuosa se rechaza el lote; en otro caso, se acepta. La calidad garantizada por el proveedor es un 8 por mil de defectuosas. Calcular la probabilidad de:
a) Aceptar un lote que contenga un 2% de defectuosas.
b) Rechazar un lote que debería ser aceptado al tener sólo el 8 por mil defectuosas.

5.1.11. Se admite que las retribuciones percibidas en una empresa se distribuyen normalmente. Se conoce, por las relaciones de seguros sociales, que el 1% son superiores a 58.000 euros y el 10% inferiores a 12.000 euros. Se pregunta qué proporción de las retribuciones son superiores a 30.000 euros.

5.1.12. En cierta fabricación mecánica el 96% de las piezas resultan con longitudes admisibles (dentro de las tolerancias), un 3% defectuosas cortas y un 1% defectuosas largas. Calcular la probabilidad de que:
a) En un lote de 250 piezas sean admisibles 242 o más.
b) En un lote de 500 sean cortas 10 o menos.
c) En 1.000 piezas haya entre 6 y 12 largas. Todas las aproximaciones se calculan mediante la distribución normal.

5.1.13. La dimensión principal de ciertas piezas tiene una distribución normal (150; 0,4) y el intervalo de tolerancia es (149,2; 150,4). Se pide:
a) La proporción esperada de defectuosas resultantes de dicho proceso.
b) Se toman 50 piezas, calcular la probabilidad de que 44 sean aceptables.

5.1.14. Si x es normal (0, 1) y se define $y = 2x^2 - 1$, calcular la probabilidad de que y no se aparte de su media más de una desviación típica.

5.1.15. Los logaritmos decimales de ciertas magnitudes, y, siguen una distribución normal. Se pide:
a) Escribir la función de densidad de $y = 10^x$, donde x es $N(2; 0{,}8)$.
b) Calcular $p(y < 15)$, $p(15 < y < 4.000)$ y $p(y > 4.000)$.
c) Calcular la probabilidad de que al tomar 10 valores al azar de y, 3 sean menores que su mediana y 4 superiores a ella.

5.1.16. Una compañía aérea, observando que, en promedio, el 12% de las plazas reservadas no se cubren, decide aceptar reservas por un 10% más de las plazas disponibles en aviones de 450 plazas. Calcular la proporción de vuelos en que algún pasajero con reserva no tiene plaza (indicar las hipótesis hechas para resolver el problema).

5.1.17. La vida (en horas) de ciertos tubos electrónicos tienen una densidad $f(x) = 0$, $x < 200$; $f(x) = ke^{-x^2/80.000}$ si $x \geq 200$ (normal truncada). Un aparato contiene 100 de estos tubos y para su funcionamiento al menos 65 de los tubos deben estar activos. Calcular la probabilidad de que el aparato funcione después de 250 horas de servicio.

5.1.18. En las observaciones de Rutherford y Geiger una sustancia radiactiva emite 3,87 partículas a cada 7,5 segundos. Calcular la probabilidad de que se emita al menos una partícula en un segundo.

5.1.19. En un proceso de fabricación la probabilidad de pieza defectuosa es p y los defectos se producen de acuerdo con el proceso de Bernoulli. Se considera la variable y número de piezas hasta la primera defectuosa (distribución de Pascal). Se pide:
a) Obtener su distribución de probabilidad.
b) Demostrar que la esperanza de y es $1/p$.
c) Con ayuda del cuadro 5.2 obtener su varianza, coeficiente de asimetría y apuntamiento.

5.1.20. En el proceso anterior consideremos la variable número de piezas totales antes de la k-ésima defectuosa. Se pide:
a) Obtener su distribución de probabilidad.
b) Obtener su esperanza.
c) Con ayuda del cuadro 5.2 obtener su varianza.

5.1.21. Justificar que la misma relación que existe entre las medidas del cuadro 3.2 para la binomial y Poisson se manifiesta entre la geométrica y la exponencial. Utilizando esta similitud, razónese cuáles serían estas medidas para el equivalente continuo de la binomial negativa.

5.1.22. Se considera un experimento con dos resultados posibles, A y $\bar{A}$, pero donde la probabilidad de A varía de experiencia en experiencia y toma en la experiencia i el valor p_i. Demostrar que la variable y número de A en n experiencias tiene esperanza $n\bar{p}$ y varianza $n\bar{p}\bar{q} - n\sigma_p^2$, siendo $\bar{p} = \Sigma pi/n$ la media y $\sigma_p^2 = 1/n\Sigma(p_i - - \bar{p})_2$ la varianza de estas probabilidades.

5.1.23. Se analizan muestras de tamaño 20 de distintos lotes de piezas y se cuentan el número de defectuosas. Si los lotes están fabricados por distintas máquinas y las probabilidades de defecto en cada lote son p_1, ..., p_n, demostrar que la varianza de la variable y número de defectos en las muestras anteriores tendrá mayor varianza que la distribución binomial. Tómese $\bar{p}$ como la probabilidad media en todos los lotes y llámese σ_p^2 a la varianza de p entre lotes.

5.7 Deducción de distribuciones: el método de Montecarlo

5.7.1 Introducción

Un problema frecuente es encontrar la distribución de probabilidad de una variable aleatoria que es una función general de otras variables conocidas. Por ejemplo, estamos interesados en conocer la distribución del tiempo que se tarde en realizar una actividad, y, que puede descomponerse en dos etapas, y conocemos la distribución del tiempo de cada etapa, x_1, x_2, donde:

$$y = x_1 + x_2 \tag{5.22}$$

Además suponemos que el tiempo invertido en la segunda etapa no depende del tiempo invertido en la primera, es decir, que ambas actividades se realizan independientemente. Diremos entonces que las variables x_1 y x_2 son independientes, un concepto que estudiaremos con más detalle en el capítulo siguiente. El problema es obtener la distribución de y.

En esta sección vamos a estudiar un método para resolver el problema anterior con ayuda del ordenador. La importancia de este método es que es completamente general y permite resolver el problema global siguiente: dadas las variables aleatorias independientes x_1, ..., x_n con distribución conocida, obtener la distribución de probabilidad de la nueva variable aleatoria unidimensional:

$$y = g(x_1, ..., x_n) \tag{5.23}$$

donde la función g es conocida.

Si en este problema tomamos $n = 2$ y como g la función suma, volvemos a (5.22). Cuando las variables x_1 y x_2 son discretas, su suma también lo será y su distribución puede obtenerse directamente combinando las dos variables de todas las formas posibles, como indica la tabla 5.2. Cuando el número de valores posibles de las variables es muy grande y cuando n sea alto, será necesario efectuar los cálculos con un ordenador, pero el procedimiento es el mismo.

Tabla 5.2 Cálculo de la distribución de una suma de variables discretas

x_1		x_2		Generación de y
1	0,5	0	0,6	(1 0) → 1 con $p = 0,5 \cdot 0,6 = 0,30$
2	0,3	1	0,4	(1 1) → 2 con $p = 0,5 \cdot 0,4 = 0,20$
3	0,2		1,0	(2 0) → 2 con $p = 0,3 \cdot 0,6 = 0,18$
	1,0			(2 1) → 3 con $p = 0,3 \cdot 0,4 = 0,12$
				(3 0) → 3 con $p = 0,2 \cdot 0,6 = 0,12$
				(3 1) → 4 con $p = 0,2 \cdot 0,4 = 0,08$

Distribución de $y = x_1 + x_2$	
1	0,30
2	0,38
3	0,24
4	0,08
	1,00

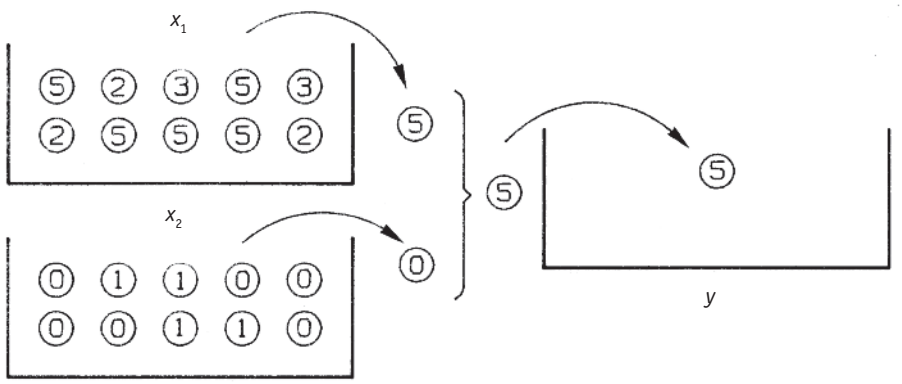
Cuando las variables x_1 y x_2 son continuas, el procedimiento anterior podría todavía aplicarse de manera aproximada convirtiéndolas en discretas (lo que requiere dividir su rango en clases), pero resulta más engorroso. Además, este procedimiento resulta difícil de generalizar para situaciones complicadas con funciones cualesquiera y variables dependientes.

Un procedimiento alternativo que siempre puede aplicarse es generar directamente la distribución de y. Para concretar, supongamos la situación de la tabla 5.2. Las etapas de resolución del problema son las siguientes:

(1) construir físicamente la variable aleatoria x_1 introduciendo en una urna 10 bolas, 5 marcadas con un uno, 3 con dos y 2 con tres;

(2) construir la variable x_2 con otra urna con 6 bolas marcadas con cero y 4 con un uno;
(3) generar un valor al alzar de cada variable extrayendo una bola de cada urna;
(4) sumar los dos valores anteriores y obtener el valor de y, que se introducirá en una tercera urna;
(5) reemplazar las bolas de x_1 y x_2 a sus urnas y repetir los pasos (3) y (4) muchas veces. Después de un número grande de repeticiones (por ejemplo 10.000), la tercera urna contendrá la distribución de la variable suma. La figura 5.9 ilustra este procedimiento:

Figura 5.9 Generación de la variable suma de x_1 y x_2



Para realizar este proceso con un ordenador, es necesario disponer de un procedimiento que simule las extracciones de las urnas, es decir, que proporciones valores al azar de una distribución conocida. El método de Montecarlo resuelve este problema.

5.7.2 El método de Montecarlo

El método de Montecarlo es un procedimiento general para seleccionar muestras aleatorias de una población (finita o infinita) de la que se conoce su distribución de probabilidad mediante números aleatorios. Se llama *números aleatorios* a conjuntos de números construidos de manera que todos los dígitos tienen la misma probabilidad de aparición. Por ejemplo, las primeras tablas de números aleatorios construidas en España se hicieron escribiendo en secuencia los números premiados en la lotería nacional en los últimos años (véase la tabla 1 en el apéndice de tablas). Los ordenadores, e

incluso algunas calculadoras de bolsillo, generan por operaciones aritméticas números pseudoaleatorios a partir de un valor inicial que se toma como semilla. Aunque estos números no son exactamente aleatorios, ya que quedan determinados por la semilla, verifican la propiedad de equiprobabilidad de aparición de cada dígito, y se utilizan mucho en la práctica.

Comencemos estudiando cómo generar valores al azar de la distribución de la variable aleatoria discreta de la tabla 5.3.

Tabla 5.3 Distribución de una variable discreta

x	*P(x)*	*F(x)*
0	0,41	0,41
1	0,26	0,67
2	0,18	0,85
3	0,10	0,95
4	0,05	1

Podemos utilizar los números aleatorios de la forma siguiente: partiendo de números de dos dígitos, los convertiremos en decimales de manera que $0 \leq NA < 1$ y estableceremos la correspondencia de la tabla 5.4 (por ejemplo, si el número aleatorio es 0,69, diremos que hemos observado $x = 2$) y tomaremos tantos números aleatorios como elementos deba contener la muestra.

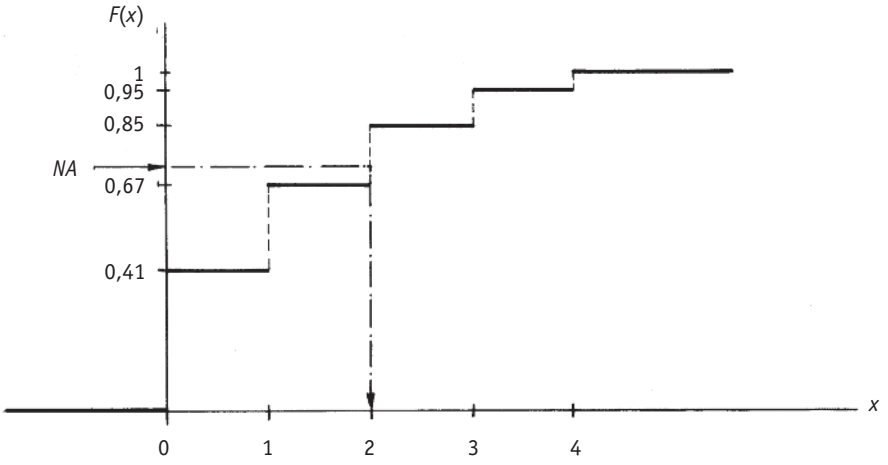
Tabla 5.4 Método de Montecarlo para la distribución de la tabla 5.3

números aleatorios (*NA*) entre	**equivalen al valor de *x***	**valor de *F*(*x*)**
0,00 - 0,40	0	0,41
0,41 - 0,66	1	0,67
0,67 - 0,84	2	0,85
0,85 - 0,94	3	0,95
0,95 - 0,99	4	1

Como hemos asignado el 41% de los números aleatorios al cero, el 26% al uno, etc., aseguramos que los valores de x van a aparecer en la muestra con sus probabilidades en la población.

Este procedimiento equivale a considerar el valor anterior, *NA,* como un valor de la función de distribución de la variable que simulamos, y tomar como observación el valor x más pequeño que verifica $F(x) > NA$. La figura 5.10 ilustra el método.

Figura 5.10 Generalización de un valor al azar de x con distribución $F(x)$ discreta



El método de la función inversa para variables continuas

El procedimiento anterior puede generalizarse para distribuciones continuas como sigue:

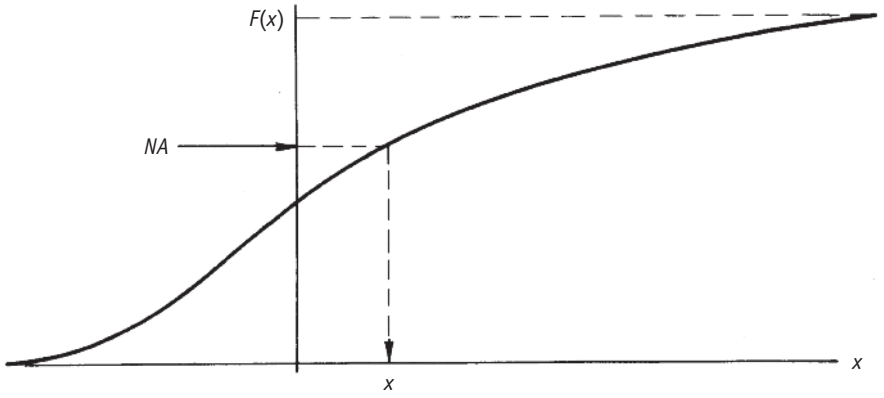
1) Tomar un número aleatorio de tantas cifras como precisión se desee y convertirlo en decimal (ej. el n.º 23.457 se convierte en 0,23457). Sea *NA* dicho valor.
2) Considerar el valor *NA* como un valor de $F(x)$ y tomar como valor observado en la muestra aquel valor x tal que $NA = F(x)$; $x = F^{-1}(NA)$.
3) Generar una muestra de tamaño n, repitiendo (1) y (2) n veces con distintos números aleatorios.

La figura 5.11 ilustra este procedimiento.

Vamos a demostrar que este procedimiento proporciona valores al azar de la variable x. Para ello veremos que los números así generados tienen precisamente la misma distribución que x. Además, serán independientes, por serlo los números aleatorios. Supondremos que los números aleatorios se toman con muchas cifras, de manera que puedan considerarse como valores al azar de una distribución uniforme en el intervalo (0, 1). Entonces, llamando u a estos números aleatorios, su función de distribución será:

$$F_u(u_0) = P(u \leq u_0) = u_0 \qquad 0 < u_0 < 1$$

Figura 5.11 Generación de un valor al azar de una distribución continua



El procedimiento que hemos expuesto consiste en generar un valor muestral y de una variable x con distribución $F_x(x)$ mediante:

$$y = F_x^{-1}(u) \tag{5.24}$$

Vamos a demostrar que la función de distribución de esta variable y, que llamaremos en general F_y, es precisamente F_x. En efecto:

$$F_y(y_0) = P(y \le y_0) = P[F_x^{-1}(u) \le y_0] = P[u \le F_x(y_0)] = F_x(y_0) \tag{5.25}$$

y, por tanto, como esto es válido para cualquier punto, $F_y = F_x$.

Este método requiere conocer la inversa de la función de distribución.

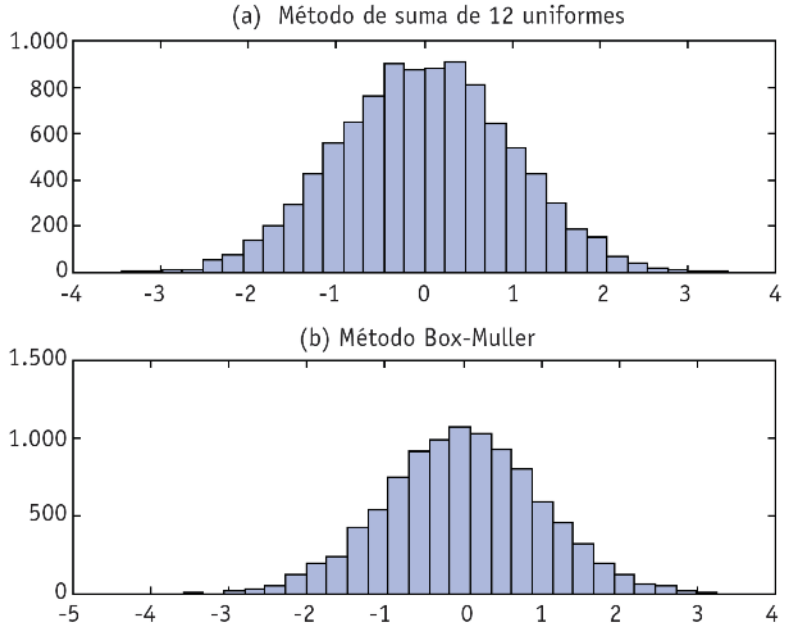
5.7.3 Aplicaciones

Vamos a comentar brevemente cómo obtener valores aleatorios al azar de las distribuciones estudiadas. La forma más inmediata de obtener valores al azar de las distribuciones ligadas a procesos binomiales es simular el proceso binomial. Sea p la probabilidad de éxito. Definimos números aleatorios con tantos dígitos como tenga p y si $NA \le p$ suponemos que $x = 1$ y cero en otro caso. Si generamos bloques de n números aleatorios y anotamos el número de unos, tenemos una muestra de la binomial, si contamos el número de unos hasta el primer cero de la geométrica, etc.

Para obtener muestras del proceso de Poisson lo más rápido es utilizar que los intervalos entre sucesos son exponenciales y utilizar la función de distribución de la exponencial, que es

$$F(x) = 1 - e^{-\lambda x}$$

Figura 5.12 Un ejemplo de ilustración de dos métodos de generación de variables normales



Como en esta función es fácil calcular la inversa, llamando u al valor aleatorio, tenemos:

$$u = 1 - e^{-\lambda x}$$

que corresponde a un valor de la variable

$$x = -\frac{1}{\lambda}\ln(1-u)$$

Para obtener el número de sucesos de Poisson en un tiempo T generamos valores de x hasta que su suma sea mayor que T. Si para ello hay que generar $k + 1$ variables, el número de sucesos de Poisson en el intervalo es k.

La generación de valores normales no puede hacerse con el método de la función inversa, pero podemos utilizar el teorema central del límite. Si generamos 12 números uniformes y definimos

$$x = u_1 + ... + u_{12} - 6$$

esta variable tiene media cero [ya que $E(u) = 0, 5$], varianza unidad [ya que para valores uniformes $Var(u) = 1/12$, como puede comprobar el lector] y

distribución normal. Este procedimiento funciona razonablemente bien, aunque la aproximación mejora tomando un mayor número de variables uniformes, con lo que se convierte en lento. Un procedimiento menos intuitivo pero más rápido es debido a Box y Muller, que demostraron que si generamos dos valores uniformes, u_1, u_2 la variable

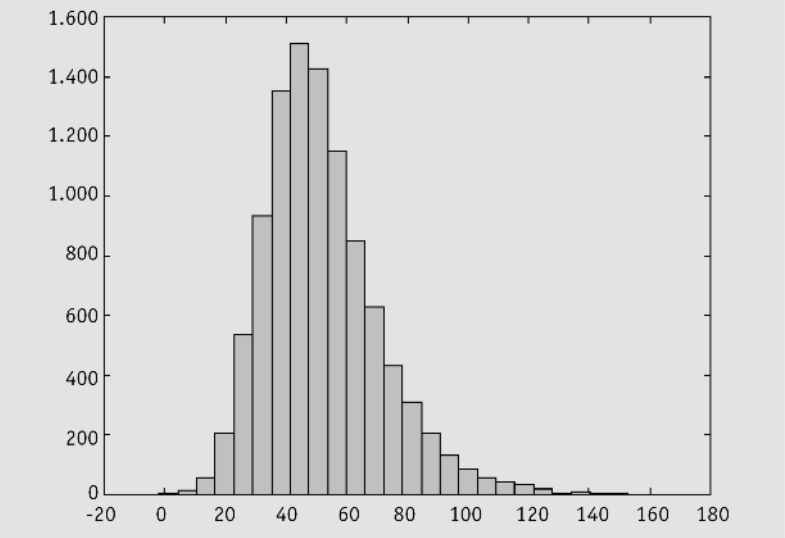
$$x = \sqrt{-2\log u_1}\cos(2\pi u_2)$$

tiene una distribución casi idéntica a la normal estándar. La figura 5.12 presenta una muestra de 10.000 observaciones generadas con cada uno de estos dos procedimientos.

Ejemplo 5.9

Un proyecto requiere tres etapas. La primera es de recogida de información, y su duración sigue una distribución exponencial con media 15 días (luego λ_1 = 1/15); la segunda, de realización, tiene una duración que sigue una variable normal de media 30 días y desviación típica 10; la tercera, redacción, es de nuevo exponencial con media 7 días, (λ3 = 1/15). Se desea calcular la distribución del tiempo total en realizar el proyecto.

Figura 5.13 Tiempo de realización de un proyecto



Una forma simple de resolver este problema es mediante el método de Montecarlo. Generaremos un valor al azar para la duración de la primera

etapa con su distribución exponencial mediante $x_1 = -15 \ln(1 - u)$ o, lo que es equivalente, mediante $x_1 = -15 \ln(u)$, ya que tan uniformes son u como $1 - u$. A continuación generaremos el tiempo de realización como $x_2 = 30 + 10z$ donde z es una variable normal estándar generada por el método de Box-Muller y obtendremos el tiempo de redacción con $x_3 = -7\ln(u)$. Repitiendo este proceso 10.000 veces con un ordenador, se obtiene el resultado de la figura 5.13.

La media obtenida para $y = x_1 + x_2 + x_3$ es 51,98, muy próxima a la suma de las medias, 52 días; la desviación típica es 19,18, el coeficiente de asimetría 1 y el de curtosis 5. Esta distribución de probabilidad nos proporciona toda la información respecto a la duración del proyecto. Por ejemplo, la probabilidad de que dure más de 100 días se obtiene con el ordenador sin más que ver la proporción de tiempos totales simulados que resultan mayores de 100, que es el 2% aproximadamente.

5.8 Distribuciones deducidas de la normal

El procedimiento anterior puede aplicarse para obtener algunas distribuciones que se deducen de la normal y que van a ser importantes en las aplicaciones. Aunque en el capítulo siguiente veremos métodos para obtener la distribución por métodos analíticos, el método de Montecarlo tiene la ventaja de la generalidad y rapidez con los ordenadores actuales.

5.8.1 La distribución χ^2 de Pearson

La distribución χ^2 fue obtenida por K. Pearson a principios del siglo XX y, según una reciente encuesta, es una de las herramientas de análisis más utilizada en la ciencia actual. Supongamos que generamos mediante el método de Montecarlo n variables aleatorias independientes normales con media cero y varianza unidad y definimos la operación:

$$\chi^2_n = z_1^2 + \ldots + z_n^2 \tag{5.26}$$

Es decir, elevamos los n valores generados al cuadrado y los sumamos. Si aplicamos este procedimiento muchas veces, obtendremos la distribución de una variable que sólo depende del número de sumandos. Esta distribución se denomina χ^2 con n grados de libertad y se representa en la figura 5.14 (su expresión matemática se indica en el cuadro 5.2).

Los parámetros de la χ^2 se obtienen fácilmente utilizando la independencia de las variables. Como $E[z_i^2] = 1$ (ya que $\sigma^2 = 1$) y $E[z_i^4] = 3$ (véase el

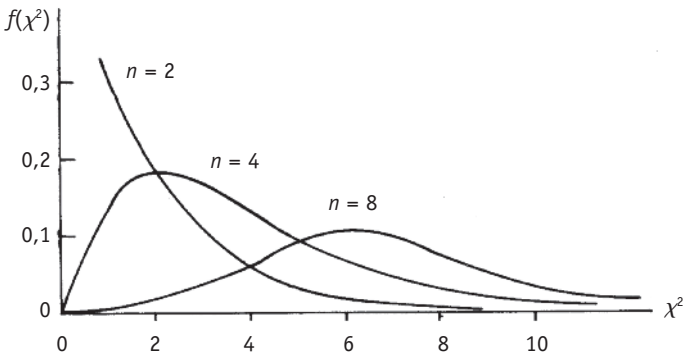
cuadro 5.2), entonces, puede comprobarse, como justificaremos en el capítulo siguiente, que:

$$E[\chi_n^2] = n \qquad Var[\chi_n^2] = 2n \tag{5.27}$$

La distribución χ^2 es asimétrica (figura 5.14) y se encuentra tabulada en función de n. Su propiedad fundamental es que si sumamos dos χ^2 independientes de grados de libertad n_1 y n_2 se obtiene una nueva variable χ^2 con grados de libertad la suma de n_1 y n_2 Esta propiedad se deduce de la definición de la variable.

La distribución χ_n^2/n representa la distribución de la varianza de n variables normales independientes. Tiene media uno y varianza $2/n$.

Figura 5.14 Distribuciones χ^2



5.8.2 La distribución *t* de Student

La distribución t fue obtenida por W. S. Gosset, un químico que trabajaba para la cervecería Guinnes en Dublín, en 1908 mediante el método de Montecarlo. Gosset buscaba un método que le permitiese juzgar si determinados tratamientos afectaban a la calidad de la cerveza y publicó su descubrimiento bajo el pseudónimo de Student porque Guinnes no permitía a sus empleados divulgar sus descubrimientos. Su expresión matemática es:

$$t_n = \frac{z}{\left(\frac{1}{n}\chi_n^2\right)^{1/2}} \tag{5.28}$$

donde z es una variable aleatoria normal estándar, independiente del denominador, y el denominador incluye la raíz cuadrada de una distribución χ^2_n dividida por sus grados de libertad. Este denominador

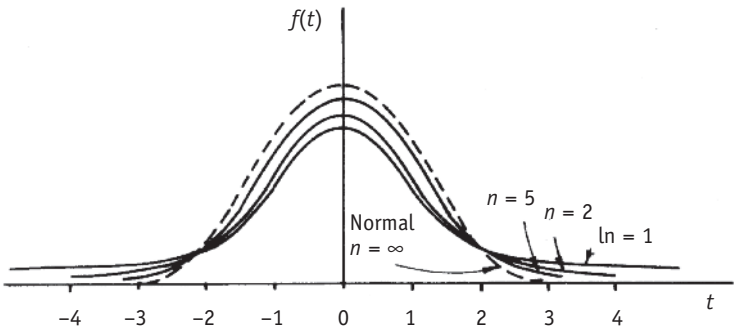
$$\left(\frac{\chi^2_n}{n}\right)^{1/2} = \sqrt{\frac{1}{n}(x_1^2 + ... + x_n^2)}$$

representa la desviación típica muestral de las variables x, ya que éstas tienen media cero. Por tanto, *la distribución t es el resultado de comparar una variable de media cero con una estimación de su desviación típica construida con n datos independientes.*

La variable t es simétrica, con mayor dispersión que la distribución normal estándar, y tiende a ésta rápidamente con n, siendo sustancialmente idéntica a la normal para n igual o mayor que 100 (véase la figura 5.15). Tiene media cero y varianza (para $n > 2$):

$$Var(t) = \frac{n}{n-2}.$$

Figura 5.15 La distribución *t*



Si efectuamos una transformación lineal de la variable t:

$$T = a + bt$$

diremos que T es una *variable t generalizada* con media a y factor de escala b. La variable T tiene propiedades análogas a la t y converge, cuando n es grande, a una variable $N(a, b)$.

5.8.3 La distribución *F* de Fisher

La distribución *F* surge al comparar dos varianzas estimadas. En efecto, es el cociente entre dos distribuciones χ^2 independientes divididas por sus grados de libertad, que, como hemos visto, representan varianzas muestrales calculadas con datos normales. Su expresión es:

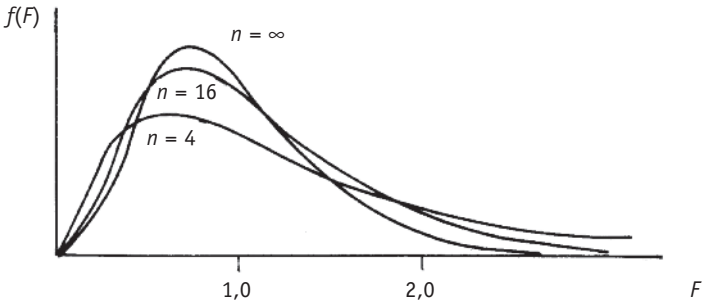
$$F_{n,m} = \frac{\dfrac{x_1^2 + \ldots + x_n^2}{n}}{\dfrac{y_1^2 + \ldots + y_m^2}{m}} = \frac{\dfrac{\chi_n^2}{n}}{\dfrac{\chi_m^2}{m}} \tag{5.29}$$

y se conoce como distribución *F* con *n* y *m* grados de libertad. La distribución se halla tabulada en función de *n* y *m*, grados de libertad del numerador y del denominador (véase la tabla 7). Por definición se verifica que $F_{n,m} = F_{n,m}^{-1}$. Algunas de sus propiedades se resumen en el cuadro 5.2.

La distribución *F* va a aparecer en la inferencia estadística al comparar varianzas de poblaciones normales. Puede considerarse una generalización de la distribución *t*, ya que se verifica la relación:

$$t_n^2 = F_{1,n} \tag{5.30}$$

Figura 5.16 La distribución *F*



5.9 Distribuciones mezcladas

Diremos que tenemos una distribución que es una mezcla de dos distribuciones cuando su función de distribución puede escribirse como

$$F(x) = (1 - \alpha)F_1(x) + \alpha F_2(x)$$

donde F_1 y F_2 son funciones de distribución y α un valor entre cero y uno que representa la probabilidad de que el elemento x provenga de la distribución F_2. Por ejemplo, si las estaturas de las mujeres siguen una distribución normal, F_1, y las estaturas de los hombres siguen una distribución normal distinta, F_2, y las mujeres son el 56% de la población y los hombres el 54%, la distribución de la estatura de una persona de esta población es

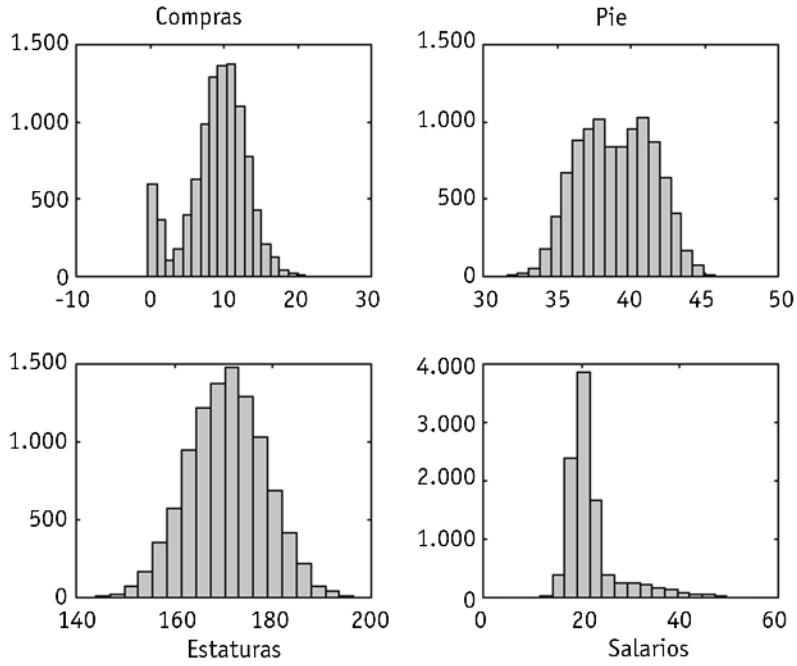
$$F(x) = 0,56F_1(x) + 0,45F_2(x)$$

Si consideramos sólo las estaturas de las mujeres, de nuevo tenemos una distribución mezclada: la distribución de las estaturas de las jóvenes será algo mayor que la distribución de las estaturas de las más mayores, y de nuevo pondríamos escribir la distribución de las estaturas de las mujeres como una distribución mezclada. Estrictamente, casi cualquier distribución que observemos en la práctica puede considerarse como una distribución mezclada, pero sólo vale la pena preocuparse por los componentes si conocerlos aumenta nuestro conocimiento de la realidad que estudiamos. En la práctica, es importante detectar que tenemos una distribución mezclada en los dos casos principales siguientes:

1. Las dos distribuciones son muy distintas y α es pequeño, es decir, la primera aparece con alta probabilidad y la segunda con baja. Esto puede ocurrir si, sin saberlo, observamos algunos datos en condiciones totalmente distintas del resto. Entonces los valores generados por la segunda distribución serán atípicos con relación a la primera, y es importante detectarlos para que no distorsionen los resultados. Por ejemplo, si cometemos errores de observación o de transcripción, etc. Estas situaciones se detectan porque la distribución mezclada tendrá un coeficiente de curtosis muy alto.
2. Las dos distribuciones están suficientemente separadas para que la distribución resultante sea bimodal. Entonces conviene identificar la causa de la discrepancia entre los datos y trabajar con las dos distribuciones. La curtosis entonces será muy baja.

Por ejemplo, la figura 5.17 presenta cuatro ejemplos de distribuciones mezcladas. Entonces la función de densidad de la distribución mezclada será $(1 - \alpha)f_1(x) + \alpha f_2(x)$, siendo f_1 y f_2 las densidades de los componentes. En el primer caso tenemos la distribución del número de compras en un centro comercial por distintos clientes. El histograma muestra que esta distribución es la mezcla de dos: los que no compran nunca o raramente, cuyo número de compras es similar a una distribución geométrica con moda en cero, y los clientes habituales, que compran una media de 10 veces en el período considerado. En el segundo caso tenemos la distribución del tamaño del calzado vendido en la sección de adultos de unos grandes almace-

Figura 5.17 Ejemplos de distribuciones mezcladas



nes. La distribución es una mezcla de las ventas a hombres y a mujeres y apuntan las dos modas de la distribución, alrededor de 38 para mujeres y 41 para hombres. En el tercer ejemplo tenemos las estaturas de un grupo de universitarios. Aunque las estaturas de las mujeres y hombres son distintas, las diferencias con relación a la variabilidad de la distribución no son grandes, y el conjunto presenta aproximadamente una distribución normal homogénea. El cuarto ejemplo corresponde a una encuesta de salarios, y vemos que aparecen dos grupos de personas: el primero, que corresponde a la mayoría de los datos, con distribución normal alrededor de 20.000 euros, que corresponde a trabajadores asalariados, y otro minoritario que corresponde a profesiones con rentas mucho mayores y más asimétricas.

Ejercicios 5.2

5.2.1. Genere 1.000 muestras de tamaño 20 de una distribución exponencial con parámetro $\lambda = 1$ por el método de Montecarlo. Calcule la media de cada muestra de 20 datos y construya un histograma para las 1.000 medias. ¿Puede explicar el resultado obtenido? Repita la generación de otras 1.000 muestras de tamaño ahora 100 y vuelva a construir el histograma; ¿qué conclusiones pueden obtenerse?

5.2.2. Obtenga una muestra de 1.000 observaciones de una distribución ji-cuadrado con 10 grados de libertad. Construya un histograma de los datos y calcule la media y la desviación típica de los 1.000 datos. Comente el resultado obtenido.

5.2.3. Obtenga una muestra de 1.000 observaciones de una distribución *t* con 3, 10 y 20 grados de libertad. Compare estas distribuciones con la normal estándar.

5.2.4. Genere por el método de Box-Muller o sumando 12 valores uniformes una muestra de tamaño 30 de una normal con media 20 y desviación típica 5. Repita el procedimiento 1.000 veces para generar 1.000 muestras de tamaño 30 y calcule la media y desviación típica de cada muestra. Haga después un histograma con las 1.000 medias y otro con las 1.000 desviaciones típicas. Comente el resultado obtenido.

5.2.5. Genere por el método de Box-Muller o sumando 12 valores uniformes una muestra de tamaño 28 de una normal con media 20 y desviación típica 5. Genere a continuación una muestra de tamaño 2 de una variable normal con media 20 y desviación típica 20. Una los 30 datos para formar una muestra de una normal mezclada, contaminada con errores de medida. Repita el procedimiento 1.000 veces para generar 1.000 muestras de tamaño 30 de esta población contaminada. Calcule la media y desviación típica de cada muestra y haga un histograma con las 1.000 medias y otro con las 1.000 desviaciones típicas. Comente el resultado obtenido.

5.10 Resumen del capítulo y consejos de cálculo

Este capítulo presenta cómo construir modelos para representar la variabilidad de una población finita o infinita. Los modelos más importantes son el binomial (para atributos), el de Poisson (para variables enteras positivas), el exponencial (para variables continuas positivas) y el modelo normal (para variables continuas cualesquiera). Sus propiedades se resumen en el cuadro 5.2, que presenta también las características de los otros modelos estudiados.

Una herramienta muy poderosa para trabajar con distribuciones de probabilidad es la generación de muestras mediante el método de Montecarlo. En particular este método puede utilizarse para comprobar la convergencia a la normal de muchas distribuciones y para obtener otras, como la χ^2, t y F, que utilizaremos en la segunda parte del libro.

En lugar de utilizar las tablas del apéndice, los programas estadísticos, incluyendo Excel, proporcionan directamente el valor de la función de distribución para los modelos estudiados. Todos permiten generar números aleatorios uniformes, pero además muchos programas proporcionan direc-

Cuadro 5.2

Nombre	Función de probabilidades o de densidad	Media	
Binomial	$p(x = r) = \binom{n}{r} p^r (1-p)^{n-r}$ $(r = 0,1, ..., n)$	np	
Geométrica	$p(x = r) = pq^{r-1}$ $(r = 1, 2, ...)$	$\frac{1}{p}$	
Poisson	$p(x = r) = \frac{\lambda^r e^{-\lambda}}{r!}$ $(r = 0,1, ...)$	λ	
Exponencial	$f(x) = \lambda e^{-\lambda x}$ $x > 0$	$1/\lambda$	
Normal	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-[(x-\mu)/\sigma]^2/2}$ $-\infty < x < \infty$	μ	
Uniforme	$f(x) = \frac{1}{b-a}$ $(a < x < b)$	$\frac{b+a}{2}$	
log normal	$f(x) = \frac{1}{x\sigma}\frac{1}{\sqrt{2\pi}} e^{-[(\ln x-\mu)/\sigma]^2/2}$ $x > 0$	$\exp\left(\mu + \frac{\sigma^2}{2}\right)$	
t con n g. de l.	$f(x) = k\left(1 + \frac{x^2}{n}\right)^{-[(n+1)/2]}$ $-\infty < x < \infty$	0	
χ^2 con n g. de l.	$f(x) = k(x^2)^{n/2-1} e^{-x^2/2}$ $x^2 \geq 0$	n	
F con n_1, n_2 g. de l.	$f(x) = kx^{n_1/2-1}$ $(n_2 + n_1 x)^{(n_1+n_2)/2}$ $x > 0$	$\frac{n_2}{n_2 - 2}$ $(n_2 > 2)$	

	Varianza	C. asimetría	C. apuntamiento
	npq	$\frac{q-p}{\sqrt{npq}}$	$3+\frac{1-6pq}{\sqrt{npq}}$
	$\frac{q}{p^2}$	$\frac{1+q}{\sqrt{q}}$	$3+\frac{p^2+6q}{q}$
	λ	$1/\sqrt{\lambda}$	$3 + 1/\lambda$
	$1/\lambda^2$	2	9
	σ^2	0	3
	$\frac{1}{12}(b-a)^2$	0	1.8
	$e^{\sigma^2}(e^{\sigma^2}-1)e^{2\mu}$	$(e^{\sigma^2}+2)\sqrt{e^{\sigma^2}-1}$	—
	$\frac{n}{n-2}$ $(n>2)$	0	$3+\frac{6}{n-4}$ $(n>4)$
	$2n$	$\sqrt{8/n}$	$3 + 12/n$
	$\frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}$ $n_2>4$	—	—

tamente valores al azar de cualquier distribución especificando sus parámetros. Recomendamos al lector obtener una muestra grande de valores de las distintas distribuciones y comparar los resultados teóricos con los observados.

5.11 Lecturas recomendadas

Los modelos aquí presentados aparecen, con mayor o menor detalle, en las referencias de cálculo de probabilidades del capítulo anterior. Johnson y Kotz (1972) es una exhaustiva recopilación de modelos de distribución de probabilidad, y Patel y Read (1996) se concentran en la distribución normal. Una clara introducción al método de Montecarlo se encuentra en el primer capítulo de Gamerman y Lopes (2006), y un estudio más extendido en Devroye (1986). Las distribuciones mezcladas se estudian en Titterington, Smith y Makov (1987) y McLachlan y Peel (2000).

Apéndice 5A: Función generatriz de momentos

Las operaciones con variables aleatorias y el cálculo de los momentos se simplifican utilizando la función generatriz de momentos. Dada una variable aleatoria x, esta función se define como:

$$\Psi(t) = E(e^{tx})$$

donde t es un número real. Por ejemplo, para la distribución exponencial

$$\Psi(t) = \int_0 e^{tx}\lambda e^{-\lambda x}dx = \frac{\lambda}{\lambda - t}$$

y verifica $\Psi(0) = 1$. Para obtener los momentos respecto al origen de la variable x, basta derivar la función generatriz y particularizar para $t = 0$. En efecto:

$$\Psi'(0) = \left[\frac{d}{dt}E[e^{tx}]\right]_{t=0} = E\left[\frac{d}{dt}e^{tx}\right]_{t=0} = E[xe^{tx}]_{t=0} = E[x]$$

en general:

$$\Psi^k(0) = \left[\frac{d^k}{dt^k}E[e^{tx}]\right]_{t=0} = E[x^k e^{tx}]_{t=0} = E[x^k]$$

Por ejemplo, para la distribución exponencial:

$$\Psi'(0) = \left[\frac{\lambda}{(\lambda - t)^2}\right]_{t=0} = \frac{1}{\lambda} \quad ; \quad \Psi''(0) = \left[\frac{+2\lambda}{(\lambda - t)^3}\right]_{t=0} = \frac{2}{\lambda^2}$$

El logaritmo de esta función se denomina función generatriz de cumulantes y sus tres primeras derivadas proporcionan respectivamente la media, varianza y momento de tercer orden respecto a la media (μ_3). En efecto, si:

$$\alpha(t) = \ln \Psi(t)$$

$$\alpha'(0) = \frac{1}{\Psi(0)} \Psi'(0) = \Psi'(0)$$

$$\alpha''(0) = \frac{\Psi''(0)\Psi(0) - [\Psi'(0)]^2}{\Psi(0)^2} = \Psi''(0) - [\Psi'(0)]^2$$

$$\alpha'''(0) = \Psi'''(0) - 3\Psi'(0)\Psi''(0) + 2\Psi'(0)^3$$

Vamos a aplicar estos resultados para obtener las medias y varianzas de las distribuciones binomial, Poisson y exponencial.

Distribución binomial

La función generatriz de momentos es:

$$\Psi_B(t) = \sum_0^n e^{tx} \sum_x^n p^x q^{n-x} = \sum_0^n \sum_x^n (pe^t)^x q^{n-x} = (pe^t + q)^n$$

y la función generatriz de cumulantes:

$$\alpha_B(t) = n \ln(pe^t + q)$$

$$\mu = \alpha'_B(0) = \left[\frac{npe^t}{pe^t + q}\right]_{t=0} = np$$

$$\sigma^2 = \alpha''_B(0) = \left[\frac{npe^t(pe^t + q) - np^2e^{2t}}{(pe^t + q)^2}\right]_{t=0} = npq$$

Distribución de Poisson

$$\Psi_p(t) = \sum_0^\infty e^{tx} \frac{(e^{-\lambda}\lambda^x)}{x!} = e^{-\lambda} \sum_0^\infty \frac{(\lambda e^t)^x}{x!} = e^{\lambda} e^{\lambda e^t}$$

$$\alpha_p(t) = \lambda(1 + e^t)$$

$$\alpha'_p(0) = [\lambda e^t]_{t=0} = \lambda \quad ; \quad \alpha''_p(0) = [\lambda e^t]_{t=0} = \lambda$$

Distribución exponencial

Según hemos visto:

$$\alpha_t(t) = \ln \lambda - \ln (\lambda - t)$$

$$\alpha'_p(0) = \left[\frac{1}{\lambda - t}\right]_{t=0} = \frac{1}{\lambda} \; ; \; \alpha''_p(0) = \left[\frac{1}{(\lambda - t)^2}\right] = \frac{1}{\lambda^2}$$

Este procedimiento simplifica en general el cálculo de momentos, que siempre puede efectuarse directamente aplicando la definición, pero suele ser más laborioso. Por ejemplo, para una variable exponencial el cálculo directo es:

$$E[t] = \int_0^\infty \lambda t e^{-\lambda t} dt$$

Llamando

$$\lambda t = u \; ; \; e^{-\lambda t} dt = dv$$

entonces

$$du = \lambda dt; \quad v = -\frac{1}{\lambda} e^{-\lambda t}$$

y como:

$$\int u dv = uv - \int v du$$

$$\int_0^\infty \lambda t e^{-\lambda t} dt = -te^{-\lambda t}\Big]_0^\infty + \int_0^\infty e^{-\lambda t} dt$$

como x/e^x tiende a cero cuando x tiende a infinito (como puede comprobarse desarrollando e^x en serie de Taylor), tendremos que

$$E[t] = \int_0^\infty \lambda t e^{-\lambda t} dt = \int_0^\infty e^{-\lambda t}\, dt = \frac{e^{-\lambda t}}{\lambda}\Bigg]_0^\infty = \frac{1}{\lambda}$$

La varianza se calcula análogamente, integrando por partes.

Aplicación a la suma de variables

Una propiedad fundamental de la función generatriz es que si $y = x_1 + ... + + x_n$ donde las x son independientes entre sí, la función generatriz de y es el producto de las funciones generatrices de las x. En efecto:

$$\Psi_y(t) = E[e^{ty}] = E[e^{tx_1} \cdot e^{tx_2}, ..., e^{tx_n}] = \Psi_{x_1}(t) ... \Psi_{x_n}(t)$$

En consecuencia, la función generatriz de cumulantes será la suma de las funciones generatrices de los sumandos. Por tanto, si sumamos, por ejemplo, varias variables de Poisson con parámetros $\lambda_1, ..., \lambda_n$, la función generatriz de cumulantes de la suma será:

$$a_y(t) = \Sigma\alpha_{x_i}(t) = (\lambda_1 + ... + \lambda_n)(1 + e^t) = \lambda_T(1 + e^t)$$

que es una distribución de Poisson con parámetros $\lambda_T = \Sigma\lambda_i$. Este procedimiento es el que se utiliza para demostrar que la suma de variables independientes binomiales es binomial, la suma de normales independientes normal, etc.

Apéndice 5B: Distribución hipergeométrica

La distribución hipergeométrica es la equivalente a la binomial, pero cuando el muestreo se hace sin reemplazamiento. Suponemos una población de tamaño N donde hay Np elementos A y Nq, $(q = 1 - p)$, elementos D. La variable hipergeométrica es el número de elementos A en una muestra de tamaño n, de donde:

$$P(x) = \frac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}}$$

sus parámetros son:

$$E[x] = np \quad ; \quad \text{Var}[x] = npq\,\frac{N-n}{N-1}$$

cuando $N \to \infty$, la distribución hipergeométrica coincide con la binomial.

Apéndice 5C: Distribución gamma

Si consideramos en un experimento de Poisson la variable X = tiempo que transcurre hasta la ocurrencia del r-ésimo éxito, su distribución será:

$$P(x > t) = P \text{ (menos de } r \text{ sucesos en } t) = \sum_{0}^{r-1} e^{-\lambda t}\,\frac{(\lambda t)^2}{x!}$$

Por tanto:

$$F(t) = P\,(x \le t) = 1 - \sum_{0}^{r-1} e^{-\lambda t}\,\frac{(\lambda t)^2}{x!}$$

Derivando respecto a t para obtener la función de densidad se obtiene:

$$f(t) = \sum_{0}^{r-1} \frac{1}{(x-1)!}\,(\lambda t)^{x-1}\lambda e^{-\lambda t} + \sum_{0}^{r-1} \frac{(\lambda t)^x e^{-\lambda t}}{x!} \cdot \lambda =$$

$$= \frac{1}{(r-1)!}\,\lambda^r t^{r-1} e^{-\lambda t} \qquad\qquad (t > 0)$$

Si generalizamos esta función para cualquier valor de r positivo, aunque no necesariamente entero, la función de densidad resultante se denomina función gamma. Cuando $r = 1$ se obtiene la distribución exponencial.

Se comprueba que:

$$E[t] = \frac{r}{\lambda} \quad ; \quad \text{Var}[t] = \frac{r}{\lambda^2} \quad ; \quad CA = \frac{2}{\sqrt{r}} \quad ; \quad CA_p = 3 + \frac{6}{r}$$

La distribución χ^2 de Pearson es un caso particular de la gamma con $\lambda = 1/2$ y $n = 2r$.

Apéndice 5D: Distribución beta

La distribución beta aparece en la estimación bayesiana en el capítulo 4. Se dice que una variable sigue la distribución beta si su función de densidad es

$$f(x) = kx^r(1-x)^{n-r} \qquad 0 < x < 1;\ r > -1;\ n > r - 1$$

la constante k es $\Gamma(n+2)/\Gamma(r+1)\Gamma(n-r+1)$. Se demuestra que los momentos son:

$$E[x] = \frac{r+1}{n+2} \ ;\ \mathrm{Var}(x) = \frac{(r+1)(n-r+1)}{(n+2)^2(n+3)}$$

La moda de la distribución es r/n. La distribución es simétrica si $r = n/2$, y asimétrica en caso contrario. Para $r = n = 0$ la distribución se reduce a la uniforme. Al aumentar r y n la distribución se va concentrando alrededor de la moda.

6. Modelos multivariantes

Harold Hotelling (1895-1973)
Científico estadounidense. Creador de procedimientos para el análisis multivariante de datos que son de uso común en todas las ramas de la ciencia. Hizo también contribuciones fundamentales a la teoría económica. Fue el creador de Statistical Research Group en la Universidad de Columbia en Nueva York y del Departamento de Estadística en Chapel Hill, uno de los centros líderes en la investigación estadística moderna.

6.1 Variables aleatorias vectoriales

6.1.1 Concepto

Cuando en lugar de observar una característica numérica (o que convertimos en numérica definiendo una variable aleatoria) observamos *n* características en cada elemento de una población, diremos que se dispone de una *variable aleatoria vectorial* o *multidimensional.* Por ejemplo, al medir la estatura, el peso y la edad en la población española, resulta una variable aleatoria tridimensional; la población y la renta de los países del mundo formarán una variable aleatoria bidimensional, etc.

En estas situaciones cada valor de la variable aleatoria es un conjunto de *n* valores numéricos. Diremos que se ha definido la *distribución conjunta* de la variable aleatoria cuando se especifique:

a) El espacio muestral o conjunto de valores posibles. Representando cada valor por un punto en el espacio de dimensión *n*, el espacio muestral es, en general, un subconjunto del espacio *n*-dimensional de los números reales.
b) Las probabilidades de cada posible resultado (subconjunto de puntos) del espacio muestral.

Diremos que la variable vectorial *n*-dimensional **X** es *discreta* si cada una de las *n*-variables escalares que la componen es discreta. Análogamente, **X** será *continua* si sus componentes lo son. Cuando algunos de sus componentes son discretos y otros continuos, diremos que la variable vectorial es *mixta.*

6.1.2 Distribución conjunta

Dada una variable aleatoria vectorial discreta, que supondremos para simplificar bidimensional, definiremos su distribución de probabilidad mediante la *función de probabilidad conjunta* $p(x_1, x_2)$, que proporciona las probabilidades de cada posible valor. Como en el caso unidimensional, esta función deberá verificar:

a) $p(\mathbf{X}_i) = p(x_{1i}, x_{2i}) \geq 0 \qquad \forall i$

b) $\sum_{i=0}^{\infty} p(\mathbf{X}_i) = \sum_{i=1}^{\infty} p(x_{1i}, x_{2i}) = 1$

Cuando la variable sea continua, las probabilidades vendrán determinadas por la *función de densidad conjunta*, que verifica:

a) $f(\mathbf{X}) = f(x_1, x_2) \geq 0.$

b) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2) dx_1 dx_2 = 1.$

Las probabilidades en el caso continuo se calcularán por integración de la forma habitual:

$$P(a < x_1 \leq b;\ c < x_2 \leq d) = \int_a^b \int_c^d f(x_1, x_2) dx_1 dx_2 \qquad (6.1)$$

6.1.3 Distribuciones marginales

Dada una variable aleatoria vectorial n-dimensional $(x_1, ..., x_n)$, llamaremos *distribución marginal* de cada componente x_i a la distribución univariante de dicho componente, es decir, a su distribución en la población considerado aisladamente. El nombre de marginal proviene de que para distribuciones discretas bivariantes definidas por una tabla de doble entrada las distribuciones marginales se obtienen en los márgenes de la tabla al sumar por filas o por columnas. En efecto, dada la distribución conjunta $p(x_1x_2)$, las distribuciones marginales se obtienen por:

$$p(x_1) = \sum_{\forall x_2} p(x_1x_2)$$

$$p(x_2) = \sum_{\forall x_1} p(x_1x_2)$$

Las distribuciones marginales se definen análogamente para variables continuas por:

$$f(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2)dx_2 \tag{6.2}$$

$$f(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2)dx_1 \tag{6.3}$$

y representan la función de densidad cuando consideramos cada variable aisladamente. En efecto, para calcular la probabilidad de que la variable x_1 pertenezca a un intervalo (a, b) podemos utilizar la distribución conjunta y escribir:

$$P(a < x_1 < b) = P(a < x_1 \leq b; -\infty < x_2 \leq \infty) = \int_a^b dx_1 \int_{-\infty}^{\infty} f(x_1, x_2)dx_2 =$$

$$= \int_a^b f(x_1)dx_1$$

que sirve de justificación a (6.2). Intuitivamente, (6.2) suma para cada valor de x_1 fijo la probabilidad de todos los pares de valores posibles (x_i, x_2) que pueden darse con dicho valor de x_1 fijo. Obviamente esta suma proporcionará la probabilidad del valor x_1.

Ejemplo 6.1

La tabla 6.1 presenta la distribución conjunta de las variables aleatorias votar a uno de cuatro posibles partidos políticos P_1, P_2, P_3 y P_4 y nivel de ingresos, A (alto), M (medio), B (bajo). Calcular las distribuciones marginales.

Tabla 6.1 Distribución conjunta de votos e ingresos en una población

	A	M	B
P_1	0,1	0,05	0,01
P_2	0,05	0,20	0,04
P_3	0,04	0,25	0,07
P_4	0,01	0,1	0,08

Para calcular la distribución marginal añadimos a la tabla una fila y una columna y colocamos allí el resultado de sumar las filas y las columnas de la tabla. Con esto se obtiene la tabla 6.2. Por ejemplo, la distribución marginal de los ingresos es: ingresos altos, probabilidad 0,2, medios, probabilidad 0,6, y bajos, 0,2.

Tabla 6.2 Distribución conjunta y marginales de votos e ingresos en una población

	A	M	B	Marginal de votos
P_1	0,1	0,05	0,01	0,16
P_2	0,05	0,20	0,04	0,29
P_3	0,04	0,25	0,07	0,36
P_4	0,01	0,1	0,08	0,19
Marginal de ingresos	0,2	0,6	0,2	1

Ejemplo 6.2

Dos amigos desayunan cada mañana en una cafetería entre las 8 y las 8,30 h. La distribución conjunta de sus tiempos de llegada es uniforme en dicho intervalo, es decir:

$$f(x, y) = k \qquad 8 \leq x \leq 8{,}30; \qquad 8 \leq y \leq 8{,}30$$
$$f(x, y) = 0 \qquad \text{en otro caso.}$$

Si los amigos acuerdan esperarse un máximo de 10 minutos, calcular la probabilidad de que se encuentren.

Para simplificar, tomemos (x, y) en minutos, es decir, $0 \leq x \leq 30$; $0 \leq y \leq \leq 30$. El espacio muestral es el cuadrado de lado 30 minutos. El suceso «encuentro» se produce si $|y - x| \leq 10$ minutos, que equivale al conjunto de puntos limitado por las rectas $y \leq x + 10$; $y \geq x - 10$ (zona rayada de la figura 6.1).

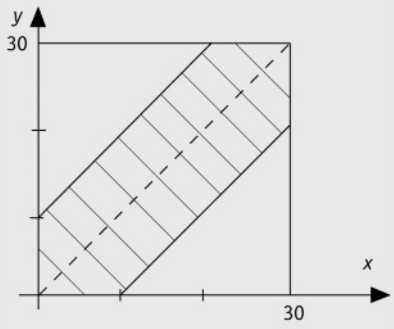
Entonces:

$$\int_0^{30} \int_0^{30} k dx dy = 1$$

que implica:

$$k = \frac{1}{900}$$

Figura 6.1 La zona rayada indica el suceso encuentro



La probabilidad de no encuentro será la integral en las dos zonas blancas triangulares de la función de densidad. Como ésta es constante:

$$\iint_A k\,dxdy = k \iint_A dxdy = k \text{ (área de } A\text{)}$$

por tanto, como el área de cada triángulo es $^1/_2\ 20 \cdot 20 = 200$:

$$P(\text{encuentro}) = 1 - P(\text{no encuentro}) = 1 - \frac{1}{900}\,2(200) = \frac{5}{9} = 0{,}55$$

La distribución marginal de *x* será:

$$\int_0^{30} k\,dy = \frac{1}{900} \cdot 30 = \frac{1}{30}$$

y es idéntica a la de *y*.

6.1.4 Distribuciones condicionadas

Se define la distribución condicionada de una variable x_1 cuando el valor de otra variable x_2 se supone fijo e igual a un valor concreto, como la distribución univariante de x_1 en los elementos de la población que tienen como valor de x_2 el valor fijado. Por ejemplo, la distribución de las estaturas condicionada a un peso de 65 kg representa la distribución de la variable estatura entre los elementos de la población estudiada que pesan 65 kg.

La distribución condicionada de una variable discreta x_1, para $x_2 = x_{20}$ fijo se obtiene «normalizando» las probabilidades conjuntas $P(x_1x_{20})$ para que sumen uno. Como:

$$\sum_{\forall x_1} P(x_1x_{20}) = P(x_{20})$$

si definimos, supuesto $P(x_{20}) \neq 0$:

$$P(x_1|x_{20}) = \frac{P(x_1x_{20})}{P(x_{20})} \tag{6.4}$$

La distribución univariante $P(x_1|x_{20})$ tendrá las propiedades de una función de probabilidades univariantes, y la llamaremos función de probabilidad de x_1 para $x_2 = x_{20}$.

La ecuación (6.4) implica que, tomando un valor genérico x_2:

$$P(x_1x_2) = P(x_1|x_2)P(x_2) \tag{6.5}$$

que relaciona las probabilidades conjuntas con las condicionadas y las marginales.

Para variables continuas, se define la distribución condicionada de x_1 para un valor concreto de la variable x_2 por:

$$f(x_1|x_2) = \frac{f(x_1, x_2)}{f(x_2)} \tag{6.6}$$

supuesto que $f(x_2) \neq 0$. Esta definición es consistente con el concepto de probabilidad condicionada y con el de función de densidad para una variable.

Ejemplo 6.3

En la distribución conjunta de los datos del ejemplo 6.1, calcular la distribución condicionada de los votos para las personas con ingresos medios y la distribución condicionada de los ingresos para los votantes del partido P_1.

Para calcular la condicionada de los votos para las personas de ingresos medios dividimos cada casilla de la columna de ingresos medios por el total de la columna. La distribución resultante se indica en la tabla 6.3.

Tabla 6.3 Distribución condicionada de los votos para personas con ingresos medios

P_1	P_2	P_3	P_4
0,0833	0,3333	0,4167	0,1667

Por ejemplo, el valor 0,0833 es el resultado de dividir 0,05, la probabilidad conjunta de ingresos medios, y votar a P_1 por la probabilidad margi-

nal de ingresos medios, 0,6. Esta tabla indica que el partido preferido para las personas de ingresos medios es el P_2 con un 41,67% de los votos, seguido del P_2 con el 33,33%.

Análogamente la tabla 6.4 indica la distribución condicionada de los ingresos para los votantes del partido P_1. El grupo más numeroso de votantes de este partido es de ingresos altos (62,5%), seguido de ingresos medios (31,25%) y bajos (6,25%).

Tabla 6.4 Distribución condicionada de los ingresos para personas que votan a P_1

A	M	B	Total
0,6250	0,3125	0,625	1

6.1.5 Teorema de Bayes

El teorema de Bayes permite responder a la cuestión siguiente: si conocemos la distribución conjunta de dos variables y hemos observado el valor de una de ellas, x_2, ¿cuál es el valor más probable de la otra? ¿Cuál es la distribución de probabilidad de la variable desconocida x_1?

Responder a estas preguntas requiere calcular la distribución condicionada $f(x_1|x_2)$ donde x_2 es un valor fijo observado. Partiendo de (6.6), el numerador puede escribirse:

$$f(x_1, x_2) = f(x_2|x_1)f(x_1) \tag{6.7}$$

mientras que el denominador, distribución marginal de x_2, puede calcularse en función de (6.3) y (6.7) como:

$$f(x_2) = \int f(x_2|x_1)f(x_1)dx_1 \tag{6.8}$$

la ecuación (6.6) puede entonces escribirse como:

$$f(x_1|x_2) = \frac{f(x_2|x_1)f(x_1)}{\int f(x_2|x_1)f(x_1)dx_1} \tag{6.9}$$

que puede interpretarse como el teorema de Bayes para funciones de densidad.

6.2 Independencia entre variables aleatorias

El concepto fundamental en el estudio conjunto de varias variables aleatorias es el concepto de *independencia*. Diremos que dos variables x_1, x_2 son independientes si el conocimiento de una de ellas no aporta información respecto a los valores de la otra: en otros términos, las propiedades de valores concretos de x_1 son las mismas cualquiera que sea el valor de x_2. Esto se expresa matemáticamente:

$$f(x_1|x_2) = f(x_1) \tag{6.10}$$

que indica que la distribución condicionada es idéntica a la marginal. En consecuencia, utilizando la definición de distribución condicionada, una definición equivalente de independencia entre dos variables aleatorias x_1, x_2 es:

$$\text{independencia si:} \quad f(x_1, x_2) = f(x_1)f(x_2) \tag{6.11}$$

es decir, dos variables aleatorias son independientes si su distribución conjunta es el producto de las distribuciones marginales. Esta definición se extiende a cualquier conjunto de variables aleatorias: diremos que las variables aleatorias $x_1, ..., x_n$, con densidad conjunta $f(x_1, ..., x_n)$, son independientes, si se verifica:

$$\boxed{f(x_1, ..., x_n) = f(x_1)f(x_2) \ldots f(x_n)} \tag{6.12}$$

La independencia conjunta es una condición muy fuerte: al ser $x_1, ..., x_n$ independientes, también lo será cualquier subconjunto de variables $(x_1 ..., x_h)$ con $h \leq n$, así como cualquier conjunto de funciones de las variables individuales $g_1(x_1), ..., g_k(x_n)$, o de conjuntos disjuntos de ellas $g_1(x_1, ..., x_i)$, $g_2(x_{i+1}, ..., x_n)$. (Compárese con la independencia de sucesos, donde es necesario exigir la condición para todos los subconjuntos de sucesos.)

Para aclarar el concepto de independencia, consideremos que en una población estudiamos las variables aleatorias estatura (y), peso (x) y cociente intelectual (z). Diremos que las variables estatura y cociente intelectual son independientes si la distribución de estaturas en personas con $z = 80$ es la misma que con $z = 100$ o con $z = 120$ e igual, en todos los casos, a la distribución de estaturas cuando miramos únicamente esta variable. Ésta es la condición (6.10). Implica que no podemos mejorar nuestra predicción de la estatura conociendo el cociente intelectual: en todos los casos su valor más probable es la moda de $f(y)$ [que es igual a $f(y|z)$ para todo valor de z].

Por el contrario, si la estatura y el peso *no* son independientes, la distribución de estaturas depende del peso y será distinta en personas de 50 kg

de peso, $f(y|x = 50)$, que en personas de 90 kg, $f(y|x = 90)$. Además estas distribuciones condicionadas serán distintas de la distribución marginal de estaturas, $f(y)$, que, según (6.8), es una media ponderada de todas ellas. En consecuencia, si tenemos que prever la estatura de una persona es informativo conocer su peso: sin conocerlo el valor más probable de la estatura es la moda de $f(y)$, pero si conocemos que su peso es igual a 70 kg, el valor más probable de la estatura es la moda de la distribución $f(y|x = 70)$.

Ejemplo 6.4

Justificar si son o no independientes las variables voto e ingresos del ejemplo 6.1.

Si las variables fuesen independientes, la distribución marginal de una variable tendría que ser igual a la distribución condicionada de esa variable dado cualquier valor de la otra. Como no es así, ya que según el ejemplo 6.3 las distribuciones condicionadas no coinciden con la marginal, las variables no son independientes.

Otra forma más larga de comprobación es calcular la distribución conjunta a partir de la hipótesis de independencia como producto de las distribuciones marginales y compararla con la distribución conjunta de las variables. Si no son iguales, las variables no son independientes. En este caso, la distribución conjunta como producto de las marginales es

Tabla 6.5 Distribución conjunta obtenida como producto de las marginales de votos e ingresos *n*

	A	M	B	Marginal de votos
P_1	0,032	0,096	0,032	0,16
P_2	0,058	0,174	0,058	0,29
P_3	0,072	0,216	0,072	0,36
P_4	0,038	0,114	0,038	0,19
Marginal de ingresos	0,2	0,6	0,2	

que es muy distinta de la dada en la tabla 6.1. Por tanto, concluimos que el voto depende del nivel de ingresos.

Ejemplo 6.5

Una junta de estudiantes está formada por diez alumnos: tres de cuarto y quinto, dos de tercero, uno de segundo y uno de primero. De los 10 alum-

nos de la junta se selecciona al azar una comisión de tres personas. Sea X el número de alumnos de cuarto en la comisión e Y el número de alumnos de quinto. Estudiar la distribución conjunta de X e Y.

Hay:

$$\binom{10}{3} = 120$$

formas distintas de elegir la comisión. Como estas 120 formas son igualmente probables, la distribución conjunta será:

$$P(x = i;\, y = j) = \frac{\binom{3}{i}\binom{3}{j}\binom{4}{3-i-j}}{\binom{10}{3}} \qquad \begin{array}{l} i = 0, 1, 2, 3 \\ j = 0, 1, 2, 3 \\ i + j \leq 3 \end{array}$$

dando valores a (i, j) se obtiene la tabla siguiente:

X \ Y	0	1	2	3	$P(x)$
0	4/120	18/120	12/120	1/120	35/120
1	18/120	36/120	9/120	0	63/120
2	12/120	9/120	0	0	21/120
3	1/120	0	0	0	1/120
$P(y)$	35/120	63/120	21/120	1/120	

Las distribuciones marginales se han indicado en los márgenes de la tabla y se calculan por:

$$P(X = i) = \sum_{j=0}^{3} P(X = i;\, Y = j)$$

La distribución condicionada de X cuando $Y = 2$ será:

$$P(X = 1 \mid Y = 2) = \frac{P(X = 1\ ;\ Y = 2)}{P(Y = 2)} = \frac{9/120}{21/120} = \frac{9}{21}$$

$$P(X = 0 \mid Y = 2) = 12/21$$

La variable aleatoria número de alumnos de cuarto en comisiones con dos alumnos de quinto toma, por tanto, únicamente los valores posibles 0, 1 con probabilidades 12/21 y 9/21. Su esperanza será:

$$E[X|Y=2] = 0 \cdot \frac{12}{21} + 1 \cdot \frac{9}{21} = 0{,}43$$

Las variables X e Y son obviamente dependientes.

Ejemplo 6.6

La variable x representa la proporción de errores de tipo A en ciertos documentos e y la proporción de errores del tipo B. Se verifica que $x + y \leq 1$ (hay otros errores posibles, C, D, etc.) y la distribución conjunta de ambas variables es:

$$f(x, y) = \begin{cases} K & 0 \leq x \leq 1;\ 0 \leq y \leq 1;\ x + y \leq 1 \\ 0 & \text{en otro caso} \end{cases}$$

Para calcular el valor de K impondremos la condición de que el volumen encerrado bajo $f(x, y)$ debe ser la unidad

$$\int_0^1 \int_0^{1-x} K\, dx\, dy = 1 = \frac{K}{2}$$

por tanto, $K = 2$. Por simetría las distribuciones marginales serán idénticas. Calcularemos la de x.

$$f(x) = \int_{-\infty}^{\infty} f(x, y)dy = \int_0^{1-x} 2\, dy = 2(1 - x)$$

La distribución condicional de x para $y = y_0$ será:

$$f(x|y_0) = \frac{f(x, y_0)}{f(y_0)} = \begin{cases} \dfrac{2}{2(1 - y_0)} = \dfrac{1}{1 - y_0} & 0 \leq x \leq 1 - y_0 \\ 0 & \text{en otro caso} \end{cases}$$

y la esperanza de la distribución condicionada:

$$E(x|y_0) = \int_0^{1-y_0} \frac{x}{1 - y_0}\, dx = \frac{1 - y_0}{2}$$

6.3 Esperanzas de vectores aleatorios

6.3.1 Concepto

La manipulación de conjuntos de variables aleatorias se simplifica utilizando la notación vectorial: dado un conjunto de variables $x_1, ..., x_n$, definiremos el vector n-dimensional $\mathbf{X}$ cuyas componentes son las variables aleatorias unidimensionales. Llamaremos *función de densidad del vector aleatorio* a la función de densidad conjunta de los componentes, y vector de medias de este vector aleatorio al vector cuyos componentes son las medias o esperanzas de sus componentes. Escribiremos el vector de medias $\boldsymbol{\mu}$ como:

$$\boldsymbol{\mu} = E[\mathbf{X}] \tag{6.13}$$

donde la esperanza operando sobre un vector o una matriz debe entenderse como el resultado de aplicar este operador (tomar medias) a cada uno de los componentes.

Funciones de variables

Generalizando esta idea, si disponemos de una función escalar $y = g(\mathbf{X})$ de un vector de variables aleatorias, el valor medio de esta función se calcula:

$$E[y] = \int yf(y)dy = \int ... \int g(\mathbf{X})f(x_1, ..., x_n)dx_1 ... dx_n \tag{6.14}$$

La primera integral tiene en cuenta que y será una variable aleatoria con una cierta función de *densidad* $f(y)$ y, por tanto, su esperanza se calcula de la forma habitual. La segunda especifica que *no* es necesario calcular $f(y)$ para determinar el valor promedio de $g(\mathbf{x})$: basta ponderar sus valores posibles por las probabilidades que dan lugar a estos valores.

Esta definición es consistente, en el sentido de que ambos términos conducen al mismo resultado.

6.3.2 Esperanza de sumas y productos

Dadas n variables aleatorias definidas conjuntamente con función de densidad $f(x_1, ..., x_n)$, se verifica:

$$\boxed{E[x_1 + x_2 + ... + x_n] = E[x_1] + ... + E[x_n]} \tag{6.15}$$

La demostración es inmediata aplicando la definición de esperanza (6.14).

Para variables independientes, como $f(x_1, ..., x_n) = f(x_1) \dots f(x_n)$, se verifica además:

$$E[x_1 \dots x_n] = \int \dots \int x_1 \dots x_n f(x_1) \dots f(x_n)\, dx_1 \dots dx_n =$$

$$= \int x_1 f_1 \int x_2 f_2 \dots \int x_1 f_1 dx_1 \int x_2 f_2 dx_2 \dots \int x_n f_n dx_n = \int x_n f_n = E[x_1]E[x_2] \dots E[x_n]$$

y la esperanza de un producto es el producto de las esperanzas.

Ejemplo 6.7

La longitud de una cola en un puesto de servicio es de 6 unidades. Si los tiempos de servicio siguen una distribución exponencial de media 5 minutos y son independientes de unas unidades a otras, calcular la media y desviación típica de la distribución de tiempo de espera para una unidad que se incorpore a la cola.

Sean x_1, ..., x_6 los tiempos de servicio de las unidades en la cola. Entonces el tiempo de espera de una nueva unidad, y, será:

$$y = x_1 + \dots + x_6$$

aplicando (6.15)

$$E(y) = E(x_1) + \dots + E(x_6) = 6 \times 5 = 30 \text{ minutos.}$$

Para calcular la desviación típica, por la independencia y (6.19)

$$DT(y) = \sqrt{Var(x_1) + \dots + Var(x_6)}$$

Como en una distribución exponencial la media y la desviación típica son iguales, tenemos que:

$$DT(y) = \sqrt{6 \times 25} = 12{,}25 \text{ minutos}$$

6.4 Covarianzas y correlaciones

6.4.1 Covarianza

La covarianza es una medida de la relación lineal entre dos variables. Se define:

$$\text{Cov}(x, y) = E[(x - \mu_x)(y - \mu_y)] = E[xy] - \mu_x \mu_y \qquad (6.16)$$

y su interpretación muestral se presentó en el capítulo 2. Se verifica que:

a) Si las variables son independientes, su covarianza es nula. En efecto, si son independientes $E[xy] = \mu_x \mu_y$. Nótese que lo contrario no es cierto. La covarianza nula no indica, en general, independencia, sino falta de relación lineal. Por ejemplo, si x es $N(0, 1)$ y definimos $y = x^2$, la covarianza entre las variables (x, y) será nula, ya que:

$$\text{Cov}(x, y) = E[x(y - 1)] = E[x^3 - x] = 0$$

y sin embargo x e y están relacionadas fuertemente aunque de manera no lineal. Veremos más adelante que si *ambas* variables tienen conjuntamente una distribución normal, la covarianza nula sí implica independencia.

b) Si modificamos la escala de medida de las variables definiendo:

$$z = ax + b$$
$$\omega = cy + d$$
$$\text{cov}\,(z, \omega) = ac\ \text{cov}\,(x, y)$$

se obtiene que la covarianza varía con las unidades de medida. Una medida adimensional de la relación lineal que elimina este inconveniente es el coeficiente de correlación.

6.4.2 Correlación

Se define el coeficiente de correlación entre dos variables (x, y) mediante:

$$\boxed{\rho(x, y) = \frac{\text{cov}\,(x, y)}{\sigma_x \sigma_y}} \qquad (6.17)$$

Se demuestra fácilmente que:

1) $|\rho(x, y)| \leq 1$.
2) *Si* $y = ax + b$, entonces $|\rho(x, y)| = 1$ (su signo es igual al de a).
3) Si las variables aleatorias son independientes, $\rho(x, y) = 0$.

6.4.3 Varianza de sumas y diferencias

Si $z = x + y$, se verifica:

$$\boxed{\text{var}\,(z) = \text{var}\,(x) + \text{var}\,(y) + 2\ \text{cov}\,(x, y)} \qquad (6.18)$$

En efecto, por definición:

$$\text{var}\,(z) = E[(x - u_x + y - \mu_y)^2]$$

Desarrollando el cuadrado y tomando esperanzas, se obtiene el resultado (6.18). Por tanto, la varianza de una suma puede ser mayor, menor o igual que la suma de varianzas, dependiendo del signo de la covarianza.

Cuando las variables son independientes, sus covarianzas son nulas y, por tanto, para variables independientes:

$$\text{Var}[x_1 + x_2 + \ldots x_n] = \text{Var}[x_1] + \ldots + \text{Var}[x_n] \tag{6.19}$$

En el caso de diferencias de variables, $z = x - y$, se comprueba fácilmente que:

$$\text{Var}(x - y) = \text{Var}(x) + \text{Var}(y) - 2\,\text{Cov}(x, y) \tag{6.20}$$

Por tanto, para variables independientes la variabilidad de $x + y$ es idéntica a la de $x - y$.

6.4.4 Matriz de varianzas y covarianzas

Llamaremos matriz de varianzas y covarianzas de un vector aleatorio x a la matriz cuadrada de orden n:

$$\boxed{\mathbf{M}_x = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})']} \tag{6.21}$$

Por tanto, llamando $\mathbf{X}' = (x_1, \ldots, x_n)$, $\boldsymbol{\mu}' = (\mu_1, \ldots, \mu_n)$, tendremos que la matriz $\mathbf{M}_x$ contiene en la diagonal las varianzas de los componentes y fuera de ella las covarianzas entre las observaciones. La matriz $\mathbf{M}_x$ será siempre simétrica y semidefinida positiva, es decir, todos los menores principales serán positivos, y dado un vector cualquiera $\boldsymbol{\omega}$ se verificará:

$$\boldsymbol{\omega}'\mathbf{M}_x\boldsymbol{\omega} \geq 0$$

Esta propiedad se comprueba definiendo una variable unidimensional por:

$$v = (\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\omega}$$

y como la varianza de v debe ser no negativa:

$$\text{var}\,(v) = E[v^2] = \boldsymbol{\omega}' E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})']\boldsymbol{\omega} \geq 0$$

Ejemplo 6.8

La distribución conjunta del número de clientes entre las 11 y las 12 de la mañana de un día laborable en dos cajas rápidas de un supermercado se indica en la tabla 6.6. Calcular el número medio de clientes en ambas cajas y la covarianza, la correlación y la matriz de varianzas y covarianzas entre ambas variables.

Tabla 6.6 Distribución de clientes en dos cajas de un supermercado

	0	1	2	3	Marginal
0	0,15	0,1	0	0	0,25
1	0,1	0,2	0,1	0	0,40
2	0	0,1	0,15	0,05	0,30
3	0	0	0,05	0	0,05
Marginal	0,25	0,40	0,30	0,05	

Si llamamos x_1 a la variable número de clientes en la primera caja y la asociamos a las filas, y x_2 al número de clientes en la segunda, y la asociamos a columnas, ambas variables tienen la misma distribución marginal. Su esperanza será

$$E(x_1) = E(x_2) = 0 \times 0{,}25 + 1 \times 0{,}4 + 2 \times 0{,}3 + 3 \times 0{,}05 = 1{,}15$$

y el número esperado de clientes en una caja es 1,15. Entre las dos cajas será

$$E(x_1 + x_2) = E(x_1) + E(x_2) = 1,\ 15 + 1{,}15 = 2{,}3$$

La covarianza será

$$cov(x_1x_2) = E(x_1x_2) - 1{,}15^2$$

y la esperanza del producto se obtiene sumando los 16 términos obtenidos multiplicando cada uno de los cuatro valores posibles de x_1 por los de x_2 y por la probabilidad conjunta que la tabla 6.6 indica para esa pareja de valores:

$$E(x_1x_2) = (0 \times 0) \times 0{,}15 + (0 \times 1) \times 0{,}1 + ... + (3 \times 2) \times 0{,}05 +$$

$$+ (3 \times 3) \times 0 = 1{,}8$$

Observemos que $E(x_1x_2) \neq E(x_1)E(x_2)$ indicando que las variables no son independientes. La covarianza es precisamente la diferencia

$$cov(x_1x_2) = 1{,}8 - 1{,}15^2 = 0{,}4775$$

Para calcular el coeficiente de correlación necesitamos las desviaciones típicas de las variables. Las varianzas son:

$$var(x_1) = var(x_2) = (0 - 1{,}15)^2 \times 0{,}25 + \ldots + (3 - 1{,}15)^2 \times 0{,}05 = 0{,}7275$$

y tendremos

$$p = \frac{0{,}4775}{0{,}7275} = 0{,}656$$

La varianza del número total de clientes entre las dos cajas será

$$var(x_1 + x_2) = var(x_1) + var(x_2) + 2cov(x_1, x_2) =$$
$$= 2 \times 0{,}7275 + 2 \times 0{,}4775 = 2{,}41$$

En resumen, la media de clientes en una caja es 1,15 con una desviación típica de $\sqrt{0{,}7275} = 0{,}85$. Para las dos cajas la media es el doble, 2,30, pero la desviación típica es $\sqrt{2{,}41} = 1{,}55$. La matriz de varianzas y covarianzas para las dos variables es

$$M = \begin{bmatrix} 0{,}7275 & 0{,}4775 \\ & \end{bmatrix}$$

6.5 Esperanzas y varianzas condicionadas

6.5.1 Esperanzas condicionadas

Se define la esperanza de una variable x_1 condicionada a otra variable aleatoria x_2 como la esperanza de la distribución de x_1 condicionado a dicho valor de x_2. Para variables discretas vendrá dada por

$$E(x_1|x_2) = \Sigma\, x_1 p(x_1|x_2)$$

donde el sumatorio está extendido a todos los posibles valores de x_1. Para variables continuas esta expresión es:

$$E(x_1|x_2) = \int x_1 f(x_1|x_2)\, dx_1$$

En general la esperanza condicionada será una función del valor x_2. Cuando x_2 es un valor fijo, la esperanza condicionada será una constante. Si x_2 es una variable aleatoria, la esperanza condicionada será también una variable aleatoria.

Existe una relación muy importante entre la esperanza de una variable y las esperanzas de la distribución de esa variable condicionada a los valores de otra. Para justificar esta relación, consideremos la distribución de la tabla 6.6 del ejemplo 6.8. La esperanza de la primera variable condicionada a que la segunda (número de clientes en la segunda caja) sea cero se calcula utilizando la primera columna de la tabla:

$$E(x_1|x_2 = 0) = 0 \times \frac{0{,}15}{0{,}25} + 1 \times \frac{0{,}1}{0{,}25} + 2 \times \frac{0}{0{,}25} + 3 \times \frac{0}{0{,}25} = 0{,}4$$

y, análogamente, se obtiene que $E(x_1|x_2 = 1) = 1$, $E(x_1|x_2 = 2) = 1{,}83$ y $E(x_1|x_2 = 3) = 2$. Estas esperanzas indican que cuando hay cero, una, dos y tres personas en una caja, el número de personas que esperamos haya en la otra es 0,4, 1, 1,83 y 2. La esperanza condicionada es una variable aleatoria que tomará estos cuatro posibles valores con probabilidades iguales a las probabilidades de 0, 1, 2 y 3 personas en la caja. El valor esperado de la esperanza condicionada es

$$0{,}4 \times 0{,}25 + 1 \times 0{,}40 + 1{,}83 \times 0{,}30 + 2 \times 0{,}05 = 1{,}15$$

que es la esperanza de la variable sin condicionar. Hemos comprobado que la esperanza de una variable puede calcularse en dos etapas como sigue: en la primera etapa calculamos todas las esperanzas de la variable condicionada a los posibles valores de otra. En la segunda ponderamos estas esperanzas por sus probabilidades de aparición, que son las probabilidades de los valores de la segunda variable. Matemáticamente podemos escribir:

$$\boxed{E(x_1) = E[E(x_1|x_2)]}$$

En el primer miembro tenemos la esperanza de la variable x_1 con relación a su distribución univariante, que es, como sabemos, idéntica a la esperanza con relación a la distribución conjunta de x_1 y x_2. En el segundo miembro esta esperanza se calcula en dos etapas. En la primera, $E(x_1|x_2)$, se calcula la esperanza de x_1 con relación a su distribución condicionada por x_2. En la segunda, se toma la esperanza del resultado respecto a la distribución de x_2. Vamos a comprobar esta expresión para variables continuas:

$$E(x_1) = \int x_1 f(x_1)dx_1 = \int\int x_1 f(x_1x_2)\, dx_1dx_2 = \int\int x_1 f(x_1|x_2)f(x_2)dx_1dx_2 =$$
$$= \int f(x_2)\, [\int x_1 f(x_1|x_2)dx_1]\, dx_2 = \int E\,[x_1|x_2]\, f(x_2)dx_2 =$$
$$= E\,[E(x_1|x_2)].$$

6.5.2 Varianzas condicionadas

La varianza de x_1 condicionada a x_2, que escribiremos $var(x_1|x_2)$, se define como la varianza de la distribución de x_1 condicionado a x_2. La varianza de la variable puede también calcularse a partir de las varianzas condicionadas. Partiendo de la identidad:

$$x_1 - \mu_1 = x_1 - E(x_1|x_2) + E(x_1/x_2) - \mu_1$$

y elevando al cuadrado y tomando esperanzas respecto a la distribución conjunta de ambas variables en ambos miembros:

$$var(x_1) = E\,[x_1 - E(x_1/x_2)]^2 + E[E(x_1/x_2) - \mu_1]^2 +$$
$$+ 2E[(x_1 - E(x_1/x_2))(E(x_1/x_2) - \mu_1)]$$

En esta expresión el doble producto, que representa la covarianza entre x_1 y $E(x_1/x_2)$, es cero. Para demostrarlo vamos a calcular la esperanza con relación a la distribución conjunta en las dos etapas que hemos visto en la sección anterior. Tomando primero la esperanza con relación a la distribución de x_1 dado x_2, el término $E(x_1/x_2) - \mu_1$ del doble producto es entonces una constante, que puede salir fuera de la esperanza, y queda $E[(x_1 - \mathrm{E}[x_1/x_2])]$ que es cero, ya que la esperanza es con relación a la distribución condicionada. Por tanto

$$E\,[(x_1 - \mathrm{E}[x_1/x_2])(E[x_1/x_2] - \mu_1)] = 0.$$

Por otro lado, como

$$E\,[E(x_1/x_2)] = E(x_1) = \mu_1,$$

el término $E[E(x_1/x_2) - \mu_1]^2$ es la esperanza de la diferencia al cuadrado entre la variable aleatoria $E(x_1/x_2)$ y su media μ_1. Por tanto:

$$\boxed{var(x_1) = E[var(x_1/x_2)] + var\,[E(x_1/x_2)]}$$

Esta expresión se conoce como *descomposición de la varianza,* ya que descompone la variabilidad de la variable en dos fuentes principales de va-

riación. Por un lado, hay variabilidad porque las varianzas de las distribuciones condicionadas, $var(x_1/x_2)$, pueden ser distintas, y el primer término promedia estas varianzas. Por otro, hay también variabilidad porque las medias de las distribuciones condicionadas pueden ser distintas, y el segundo término recoge las diferencias entre las medias condicionadas y la media total, $var[E(x_1/x_2)]$.

Observemos que la varianza de x_1 no puede ser menor que el promedio de las varianzas de las distribuciones condicionadas. En las condicionadas la variabilidad se calcula respecto a las medias condicionadas, $E(x_1/x_2)$, mientras que $var(x_1)$ mide la variabilidad respecto a la media global, μ_1. Si todas las medias condicionadas son iguales a μ_1, lo que ocurrirá por ejemplo si x_1 y x_2 son independientes, entonces el término $var[E(x_1/x_2)]$ es cero y la varianza es, exactamente, la media ponderada de las varianzas condicionadas. Si $E(x_1/x_2)$ no es constante, entonces la varianza de x_1 será mayor, y tanto más cuanto mayor sea la variabilidad de las medias condicionadas.

6.6 Transformaciones de vectores aleatorios

6.6.1 Concepto

Al trabajar con funciones de densidad de vectores aleatorios **X**, es importante recordar que, como en el caso univariante, la función de densidad tiene dimensiones. Por lo tanto, si cambiamos las unidades de medida de las variables, la función de densidad debe modificarse también. En general, si **X** es un vector de dimensión n y llamamos $f(\mathbf{X})$ a su función de densidad, y pasamos a otro vector aleatorio **Y**, de la misma dimensión, mediante la transformación uno a uno, g, definida por:

$$\begin{aligned} y_1 &= g_1(x_1, ..., x_n) \\ &\vdots \\ y_n &= g_n(x_1, ..., x_n) \end{aligned}$$

donde existen las transformaciones inversas $x_1 = h_1(y_1, ..., y_n)$, ..., $x_n = h_n(y_1, ..., y_n)$, y suponemos que todas las funciones implicadas son diferenciables, entonces puede demostrarse que la función de densidad del vector **Y** es:

$$f(\mathbf{Y}) = f(\mathbf{X}) \left| \frac{d\mathbf{X}}{d\mathbf{Y}} \right| \tag{6.22}$$

donde el término $|d\mathbf{X}/d\mathbf{Y}|$ representa el jacobiano de la transformación, dado por el determinante:

$$\left| \frac{d\mathbf{X}}{d\mathbf{Y}} \right| = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & & \vdots \\ \frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$$

que suponemos es distinto de cero en el rango de la transformación.

6.6.2 Esperanzas de transformaciones lineales

Sea $\mathbf{X}$ un vector aleatorio de dimensión n y definamos un nuevo vector aleatorio $\mathbf{Y}$ de dimensión m ($m \leq n$), con

$$\mathbf{Y} = \mathbf{AX}$$

donde $\mathbf{A}$ es una matriz rectangular de dimensiones $m \times n$. Entonces, llamando $\boldsymbol{\mu}_y$, $\boldsymbol{\mu}_x$ a sus vectores de medias y $\mathbf{M}_x$, $\mathbf{M}_y$ a las matrices de covarianza:

$$\boldsymbol{\mu}_y = \mathbf{A}\boldsymbol{\mu}_x \tag{6.23}$$

$$\mathbf{M}_y = \mathbf{A}\mathbf{M}_x\mathbf{A}' \tag{6.24}$$

donde $\mathbf{A}'$ es la matriz transpuesta de $\mathbf{A}$.

En efecto, aplicando la definición:

$$E[\mathbf{Y}] = \mathbf{A}E[\mathbf{X}]$$
$$\mathbf{M}_y = E[(\mathbf{Y} - \boldsymbol{\mu}_y)(\mathbf{Y} - \boldsymbol{\mu}_y)'] = E[\mathbf{A}(\mathbf{X} - \boldsymbol{\mu}_x)(\mathbf{X} - \boldsymbol{\mu}_x)'\mathbf{A}'] = \mathbf{A}\mathbf{M}_x\mathbf{A}'$$

La fórmula (6.15) es un caso particular de (6.23) tomando $\mathbf{A} = (1, ..., 1)$; en ese caso (6.24) generaliza (6.18), y (6.19) corresponde al caso particular en que $\mathbf{M}_x$ es diagonal.

Ejemplo 6.9

Las valoraciones de los clientes de la puntualidad (x_1), rapidez (x_2) y limpieza (x_3) de un servicio de transporte tienen unas medias, en una escala de cero a diez, de 7, 8 y 8,5 respectivamente con una matriz de varianzas y covarianzas

$$M = \begin{bmatrix} 1 & 0,5 & 0,7 \\ 0,5 & 0,64 & 0,6 \\ 0,7 & 0,6 & 1,44 \end{bmatrix}$$

Se construyen dos indicadores de la calidad del servicio. El primero es el promedio de las tres puntuaciones y el segundo es la diferencia entre el promedio de la puntualidad y la limpieza. Calcular el vector de medias y la matriz de covarianzas para estos dos indicadores.

La expresión del primer indicador es

$$y_1 = \frac{x_1 + x_2 + x_3}{3}$$

y la del segundo

$$y_2 = x_1 - x_3$$

Estas dos ecuaciones pueden escribirse matricialmente

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

El vector de medias será

$$\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 7 \\ 8 \\ 8,5 \end{bmatrix} = \begin{bmatrix} 7,83 \\ -1,5 \end{bmatrix}$$

y la matriz de varianzas covarianzas

$$My = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0,5 & 0,7 \\ 0,5 & 0,64 & 0,6 \\ 0,7 & 0,6 & 1,44 \end{bmatrix} \begin{bmatrix} 1/3 & 1 \\ 1/3 & 0 \\ 1/3 & -1 \end{bmatrix} = \begin{bmatrix} 0,73 & -0,18 \\ -0,18 & 1,04 \end{bmatrix}$$

La correlación entre estos dos indicadores es muy baja ($-0,18 / \sqrt{0,73 \times 1,04} = -0,20$), lo que sugiere que los dos indicadores están bien elegidos, ya que recogen aspectos distintos de las tres variables.

6.7 La distribución multinomial

La distribución multinomial es la generalización multivariante de la distribución binomial. Suponemos un proceso estable y sin memoria que genera elementos que pueden clasificarse en k clases distintas. Por ejemplo, observamos con reemplazamiento personas al azar de una población finita y las clasificamos en k grupos según su nacimiento o elementos de un cierto proceso de fabricación que clasificamos en k clases.

Supongamos que se toma una muestra de n elementos y definimos las k variables aleatorias:

$$x_i = \text{n.º elementos en la clase } i; \quad i = 1, ..., k$$

Entonces el vector de k-variables $\mathbf{X} = (x_1, ..., x_k)$ es una variable aleatoria k-dimensional. Estrictamente podríamos definir $k - 1$ variables, ya que el valor de la última queda fijada al conocer n y los valores de las demás, ya que siempre:

$$\sum x_i = n \tag{6.25}$$

pero, por simetría, trabajaremos con las k variables. La función de probabilidad, llamando p_i a la probabilidad de cada clase, se obtiene calculando la probabilidad de observar n_1 elementos de la primera clase, n_2 de la segunda, etc., en cualquier orden. Entonces:

$$P(x_1 = n_1, ..., x_k = n_k) = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k} \tag{6.26}$$

donde $\sum n_i = n$ y $\sum p_i = 1$. En efecto, el término combinatorio tiene en cuenta las permutaciones de n elementos cuando hay $n_1, ..., n_k$ repetidos, y el segundo se deduce por la independencia de las observaciones.

Es fácil comprobar que las distribuciones marginales son binominales, con:

$$E[x_i] = np_i, \quad DT[x_i] = \sqrt{np_i(1 - p_i)} \tag{6.27}$$

Además, cualquier distribución condicionada es multinomial. Por ejemplo, la de $k - 1$ variables cuando x_k toma el valor fijo n_k es una multinomial en las $k - 1$ variables restantes con $n' = n - n_k$.

La distribución condicionada de x_1 y x_2 cuando $x_3 = n_3, ..., x_k = n_k$ es una binominal con $n' = n - n_3 - n_4 - ... - n_k$, etc.

La ecuación (6.25) implica que las variables x_i son dependientes. Para hallar las covarianzas utilizaremos que cuando $k = 2$ el coeficiente de correlación entre x_1 y x_2 debe ser -1, ya que dado x_1, x_2 queda determinada siempre como $n - x_1$, y la relación es inversa (cuando x_1 aumenta, x_2 disminuye). Entonces, por (6.17),

$$-1 = \frac{Cov(x_1 x_2)}{\sqrt{np_1p_2}\,\sqrt{np_1p_2}}$$

llamando $q_1 = 1 - p_1 = p_2$, se obtiene

$$Cov(x_1 x_2) = -np_1 p_2$$

Puede demostrarse, aplicando la definición de covarianza (6.16) y la expresión (6.26), que este resultado es general y que, para cualquier par de variables multinomiales:

$$Cov(x_i x_j) = -np_i p_j \tag{6.28}$$

La matriz de varianzas y covarianzas de una distribución multinomial es siempre singular, como consecuencia de la relación (6.26).

Ejemplo 6.10

En un proceso administrativo ciertos documentos se clasifican como: sin errores (A_1), con errores leves (A_2), con errores graves (A_3). Se ha estimado que $p_1 = P(A_1) = 0{,}7$; $p_2 = P(A_2) = 0{,}2$; $p_3 = P(A_3) = 0{,}1$, (a) si se toman tres documentos calcular la probabilidad de que haya sólo uno de la clase A_3; (b) en una muestra de siete documentos se obtienen cinco sin errores. ¿Cuál es la probabilidad de que en dicha muestra haya un documento con errores graves?

En el caso (a) los sucesos elementales posibles son, sin tener en cuenta el orden dentro de cada suceso:

$$A_1A_1A_3 \;;\; A_1A_2A_3 \;;\; A_2A_2A_3$$

y sus probabilidades serán:

$$P(x_1 = 2, x_2 = 0, x_3 = 1) = \frac{3!}{2!0!1!}\, 0{,}7^2 \cdot 0{,}2^0 \cdot 0{,}1 = 0{,}147$$

$$P(x_1 = 1, x_2 = 1, x_3 = 1) = \frac{3!}{1!1!1!}\, 0{,}7 \cdot 0{,}2 \cdot 0{,}1 = 0{,}084$$

$$P(x_1 = 0, x_2 = 2, x_3 = 1) = \frac{3!}{0!2!1!}\, 0{,}7^0 \cdot 0{,}2^2 \cdot 0{,}1 = 0{,}012$$

Luego:

$$P(x_3 = 1) = 0{,}147 + 0{,}084 + 0{,}012 = 0{,}243$$

Naturalmente este resultado puede también obtenerse considerando la binomial $(A_3\bar{A}_3)$ con probabilidades (0,9; 0,1) y:

$$P(x_3 = 1) = \binom{3}{1} 0{,}1 \cdot 0{,}9^2 = 0{,}243$$

En el segundo caso se pide $P(x_3 = 1|x_1 = 5)$. Aplicando la definición como si $n = 7$, entonces $x_2 = 7 - 1 - 5 = 1$:

$$P(x_2 = 1, x_3 = 1|x_1 = 5) = \frac{P(x_2 = 1, x_3 = 1, x_1 = 5)}{P(x_1 = 5)}$$

$$= \frac{\dfrac{7!}{5!1!1!} 0{,}7^5 \cdot 0{,}2 \cdot 0{,}1}{\binom{7}{5} 0{,}7^5 \cdot 0{,}3^2} = \frac{2!}{1!1!} \frac{0{,}2 \cdot 0{,}1}{0{,}3^2} = 0{,}444$$

También llegamos a este resultado si tenemos en cuenta que si $x_1 = 5$, queda una binomial en x_2x_3. Cuando sólo puede ocurrir A_2 y A_3, sus probabilidades serán, llamando $\bar{A}_1$ al suceso A_1 no ocurre:

$$P(A_2|\bar{A}_1) = \frac{P(A_2\bar{A}_1)}{P(\bar{A}_1)} = \frac{P(A_2)}{P(\bar{A}_1)} = \frac{p_2}{1 - p_1} = \frac{0{,}2}{0{,}3}$$

y análogamente $P(A_3) = 0{,}1/0{,}3$. Entonces el resultado obtenido por el método anterior equivale a considerar directamente esta nueva binomial.

6.8 La normal *n*'''dimensional

Diremos que un vector aleatorio $\mathbf{X}$ sigue una distribución normal n-dimensional si su función de densidad es

$$f(\mathbf{X}) = \frac{1}{|\mathbf{M}|^{1/2}(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}(\mathbf{X} - \boldsymbol{\mu})'\mathbf{M}^{-1}(\mathbf{X} - \boldsymbol{\mu})\right\} \qquad (6.29)$$

donde $\mathbf{M}$ es la matriz de covarianzas y $\boldsymbol{\mu}$ es el vector de medias. Las propiedades principales de esta distribución son:

1) Para una variable bidimensional, la distribución tiene forma de campana, como indica la figura 6.2. Al cortar con planos perpen-

diculares al (x, y) se obtienen distribuciones normales. Por tanto, las distribuciones marginales y condicionadas son normales.

2) Para la variable n-dimensional, cualquier conjunto de $r \leq n$ variables tiene conjuntamente una distribución normal.

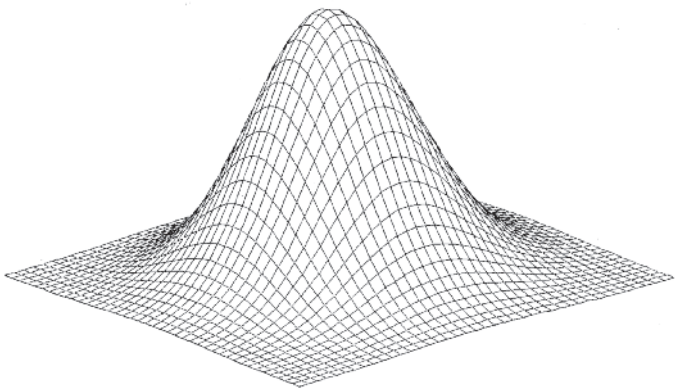
3) En la figura 6.2, al cortar por planos paralelos al plano (x, y) se obtienen las curvas de nivel representadas en la figura 6.3. Estas curvas son elipses, de ecuación:

$$[x-\mu_1\, y-\mu_2] \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}^{-1} \begin{bmatrix} x-\mu_1 \\ y-\mu_2 \end{bmatrix} =$$

$$= \left[\left(\frac{x-\mu_1}{\sigma_1} \right)^2 + \left(\frac{y-\mu_2}{\sigma_2} \right)^2 - 2\rho \left(\frac{x_1-\mu_1}{\sigma_1} \right) \left(\frac{y-\mu_2}{\sigma_2} \right) \right] \frac{1}{1-\sigma^2} = \text{cte.} \quad (6.30)$$

donde ρ representa la correlación entre las variables.

Figura 6.2 La normal bidimensional



4) Si las variables están incorreladas ($\rho = 0$), son independientes. En efecto, si **M** es diagonal, como

$$(\mathbf{X}-\boldsymbol{\mu})'\mathbf{M}^{-1}(\mathbf{X}-\boldsymbol{\mu}) = \sum \left(\frac{x_i-\mu_i}{\sigma_i} \right)^2$$

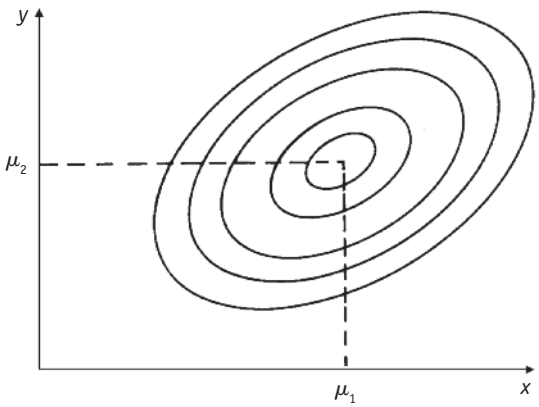
la función de densidad conjunta se descompone en el producto de las marginales. Por tanto, entre variables *conjuntamente* normales sólo pueden darse relaciones lineales.

5) Se demuestra que cualquier combinación lineal de variables aleatorias normales es también normal. Por tanto, si

$$\mathbf{Y} = \mathbf{AX}$$

donde **Y** es un vector de dimensión m ($m \leq n$), este vector tendrá una distribución normal multivariante de dimensión m, con media $\mathbf{A}\boldsymbol{\mu}_x$ y matriz de covarianzas $\mathbf{A}\mathbf{M}_x\mathbf{A}'$. La demostración es simple utilizando (6.22) para obtener la función de densidad de **Y**.

Figura 6.3 Curvas de nivel de la normal bidimensional



6) Cualquier vector **X** normal n-dimensional con matriz **M** no singular puede convertirse mediante una transformación lineal en un vector **Z** normal n-dimensional con vector de medias **O** y matriz de varianzas y covarianzas igual a la identidad **(I)**. Llamaremos normal n-dimensional estándar a la densidad de **Z**, que vendrá dada por:

$$f(\mathbf{Z}) = \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\mathbf{Z}'\mathbf{Z}\right\} =$$

$$= \prod_{i=1}^{n} \frac{1}{(2\pi)^{1/2}} \exp\left\{-\frac{1}{2}\, z_i^2\right\} \qquad (6.31)$$

La demostración es inmediata. Al ser **M** definida positiva existe una matriz cuadrada (no única) **A** que verifica:

$$\mathbf{M} = \mathbf{A}\mathbf{A}'$$

Definiendo:

$$\mathbf{Z} = \mathbf{A}^{-1}(\mathbf{X} - \boldsymbol{\mu}) \qquad (6.32)$$

tendremos que, por (6.23):

$$E[\mathbf{Z}] = \mathbf{A}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}) = 0$$

y llamando $\mathbf{M}_z$ a la matriz de varianzas de $\mathbf{Z}$, por (6.24):

$$\mathbf{M}_z = \mathbf{A}^{-1}\mathbf{M}(\mathbf{A}^{-1})' = (\mathbf{A}^{-1}\mathbf{A})(\mathbf{A}'[\mathbf{A}']^{-1}) = \mathbf{I}$$

con lo que $\mathbf{Z}$ tendrá la densidad (6.31). Por tanto, cualquier vector de variables normales $\mathbf{X}$ puede transformarse mediante (6.32) en otro vector $\mathbf{Z}$ de variables normales independientes y de varianza unidad.

Ejemplo 6.11

La distribución de dos variables (x, y) sigue una distribución normal bivariante con medias 4 y 6 y matriz de varianzas y covarianzas.

$$\begin{bmatrix} 1 & 0{,}8 \\ 0{,}8 & 2 \end{bmatrix}$$

Al analizar un elemento se observa que el valor de x es 6. ¿Cuál es el valor más probable de su valor de y?

El valor más probable para y será la media de la distribución condicionada $f(y|x = 6)$, que se obtiene por:

$$f(y|x) = \frac{f(xy)}{f(x)}$$

La distribución marginal de x es normal, $N(4,1)$. Los términos de $f(x, y)$ serán:

$$|\mathbf{M}|^{1/2} = (\sigma_1^2\sigma_2^2[1 - \rho^2])^{1/2} = \sigma_1\sigma_2\sqrt{1 - \rho^2}$$

$$\mathbf{M}^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1 - \rho^2)} \begin{vmatrix} \sigma_2^2 & -\rho\sigma_2\sigma_1 \\ -\rho\sigma_2\sigma_1 & \sigma_1^2 \end{vmatrix}$$

y el exponente de la normal bivariante $f(x, y)$ será:

$$-\frac{1}{2(1 - \rho^2)}\left\{\left(\frac{x - \mu_1}{\sigma_1}\right)^2 + \left(\frac{y - \mu_2}{\sigma_2}\right)^2 - 2\rho\frac{(x - \mu_1)(y - \mu_2)}{\sigma_1\sigma_2}\right\} = -\frac{A}{2}$$

En consecuencia, tendremos:

$$f(y|x) = \frac{(\sigma_1\sigma_2\sqrt{1-\rho^2})^{-1}\,(2\pi)^{-1}\exp\left\{\dfrac{-A}{2}\right\}}{\sigma_1^{-1}(\sqrt{2\pi})^{-1}\exp\left\{-\dfrac{1}{2}\left(\dfrac{x_1-\mu_1}{\sigma_1}\right)^2\right\}} =$$

$$= \frac{1}{\sigma_2\sqrt{1-\sigma^2}}\,\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}B\right\}$$

donde el término resultante en el exponente, que llamaremos *B*, será:

$$B = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y-\mu_2}{\sigma_2}\right)^2 - 2\rho\,\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} - \right.$$

$$\left. - \left(\frac{x_1-\mu_1}{\sigma_1}\right)^2(1-\rho^2)\right] = \frac{1}{1-\rho^2}\left[\left(\frac{y-\mu_2}{\sigma_2}\right) - \rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\right]^2$$

$$B = \frac{1}{\sigma_2^2(1-\rho^2)}\left[y - \left(\mu_2 + \rho\,\frac{\sigma_2}{\sigma_1}\,[x-\mu_1]\right)\right]^2$$

Este exponente corresponde a una distribución normal con media:

$$E[y|x] = \mu_2 + \rho\,\frac{\sigma_2}{\sigma_1}\,(x-\mu_1)$$

y desviación típica:

$$DT[y|x] = \sigma_2\sqrt{1-\rho^2}$$

Por tanto, el valor esperado de *y* aumenta linealmente con *x* según una recta que se denomina recta de regresión. Observemos que el coeficiente de correlación es la pendiente de la recta cuando estandarizamos ambas variables. Llamando:

$$Z_2 = \frac{y-\mu_2}{\sigma_2} \qquad\qquad Z_1 = \frac{y-\mu_1}{\sigma_1}$$

$$E[Z_2|x] = E[Z_2|Z_1] = \rho Z_1$$

Por ejemplo, en este caso $\rho = 0{,}8/\sqrt{2} = 0{,}57$ y la recta de regresión indica que los elementos con un *x* mayor que la media en *K* desviaciones

típicas tendrán un valor medio de *y* igual a 0,57 *K* desviaciones típicas por encima de su media. En concreto, para $x = 6$.

$$E[y|6] = 6 + \left(\frac{0{,}8}{\sqrt{2}}\right) \cdot \frac{\sqrt{2}}{1}(6 - 4) = 7{,}6 \text{ gr}$$

que es la mejor estimación del valor de *y* para $x = 6$.

Además:

$$Var[y|x] = \sigma_2^2(1 - \rho^2) = 2(1 - 0{,}32) = 1{,}36$$

que es menor que la original. Observemos que ρ^2 puede escribirse:

$$\rho^2 = \frac{\sigma_2^2 - Var(y|x)}{\sigma_2^2}$$

con lo que se interpreta como el % de reducción de varianza de la distribución que supone conocer la variable *x*. [Desconociendo el valor de *x,* la varianza de la distribución de *y* es σ^2, y al observarlo se reduce a $Var(y|x)$.]

Supongamos ahora que en lugar de trabajar con las variables originales lo hacemos con nuevas variables, *a* y *b*, definidas por:

$$a = (x + y)/2$$
$$b = x + 5y$$

y se desea obtener la distribución conjunta de *a* y *b*.

Escribiendo las relaciones anteriores como:

$$Y = \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0{,}5 & 0{,}5 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \mathbf{AX}$$

La distribución conjunta de **Y** será normal bivariante con parámetros

$$E[\mathbf{Y}] = \begin{bmatrix} 0{,}5 & 0{,}5 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} 4 \\ 6 \end{bmatrix} = \begin{bmatrix} 5 \\ 34 \end{bmatrix}$$

$$\underline{\Sigma}_y = \begin{bmatrix} 0{,}5 & 0{,}5 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} 1 & 0{,}8 \\ 0{,}8 & 2 \end{bmatrix} \begin{bmatrix} 0{,}5 & 1 \\ 0{,}5 & 5 \end{bmatrix} = \begin{bmatrix} 1{,}15 & 7{,}9 \\ 7{,}9 & 5{,}9 \end{bmatrix}$$

Ejercicios 6

6.1. La función de probabilidad de (x, y) es $p(x, y) = 1/30$ para $x = 0, 1, 2, 3, 4, 5$ e $y = 0, 1, 2, 3, 4$; $p(x, y) = 0$ en puntos distintos de los anteriores. Calcular la función de distribución en los puntos de la recta $x - 2y + 2 = 0$.

6.2. Una pareja se cita entre las 7 y las 8 de la tarde y llegan a la cita con distribución uniforme en dicho intervalo. Deciden esperarse un máximo de 15 minutos. Calcular la probabilidad de que se encuentren.

6.3. La variable bidimensional (x, y) tiene como función de densidad $f(x, y) = e^{-(x+y)}$ en el primer cuadrante; $f(x, y) = 0$ en los otros tres. Si se toman al azar tres puntos en el primer cuadrante, calcular la probabilidad de que uno al menos pertenezca al cuadrado $(0 \le x \le 1; 0 \le y \le 1)$.

6.4. En un aparato de control actúan dos variables x_1, x_2 independientes, ambas con distribución uniforme, la primera entre 1 y 9, la segunda entre 1 y a. El aparato funciona bien cuando $x_1 < 4x_2^2$. Calcular el valor de a para que $p(x_1 > 4x_2^2) \le 0{,}01$.

6.5. Dada $f(x, y) = 3x(0 < y < x; 0 < x < 1)$, obtener las distribuciones marginales y la condicionada $f(x/y)$.

6.6. El tiempo total que un camión permanece en un almacén está definido por una variable aleatoria x. Sea y la variable tiempo de espera en la cola y z el tiempo de descarga $(x = y + z)$. La distribución conjunta de x e y es:

$$f(x, y) = \begin{cases} \frac{1}{4} e^{-x/2} & 0 \le y \le x < \infty \\ 0 & \text{en otro caso} \end{cases}$$

Se pide:

a) calcular el tiempo medio total que permanece un camión en la estación;
b) calcular el tiempo medio de descarga;
c) calcular el coeficiente de correlación entre el tiempo total y el tiempo de espera en la cola.

6.7. Demostrar que la condición necesaria y suficiente para que dos variables x, y con distribución discreta definida por una tabla de valores $p(x, y)$ sean independientes es que las filas y las columnas sean proporcionales entre sí.

6.8. Si $y_1, ..., y_n$ son variables con media μ y matriz de covarianza $\sigma^2 I$, donde I es la matriz unidad, calcular la media y varianza de la variable $z = \sum a_i y_i$ y la correlación (y_i, z).

6.9. Una máquina de empacado automático deposita en cada paquete 81,5 g, por término medio, de cierto producto, con $\sigma = 8$ g. El peso medio del paquete vacío es 14,5 g, con $\sigma = 6$ g. Ambas distribuciones son normales e independientes. Se pide:
   a) calcular la distribución del peso de los paquetes llenos;
   b) escribir la distribución conjunta del peso del paquete y del producto que contiene.

6.10. En el problema 6.9 los paquetes se distribuyen en cajas de 40, cuyo peso medio vacías es 520 g, con $\sigma = 50$ g. Calcular:
   a) la distribución del peso de las cajas llenas;
   b) la probabilidad de que un cajón vacío pese menos que 5 paquetes llenos.

6.11. Una línea eléctrica se avería cuando la tensión sobrepasa la capacidad de la línea. Si la tensión es $N(100; 20)$ y la capacidad $N(140; 10)$, calcular la probabilidad de avería, suponiendo que la tensión y la capacidad varían independientemente.

6.12. Se toman tres mediciones independientes y_1, y_2, y_3 de la tensión en un circuito con tres aparatos, cuyas varianzas son 1, 2 y 3. Se forman dos índices del circuito por:

$$z_1 = 3y_1 + 2y_2 + 5y_3$$
$$z_2 = \frac{1}{3}y_1 + \frac{1}{3}y_2 + \frac{1}{3}y_3$$

Calcular el coeficiente de correlación entre z_1 y z_2.

6.9 Resumen del capítulo y consejos de cálculo

Podemos modelar la dependencia conjunta de varias variables mediante su distribución conjunta. En la práctica son muy importantes las distribuciones condicionadas de una variable dadas las demás, que nos van a resolver el problema de prever los valores de una variable conocidos los valores de otras. En particular, la esperanza condicionada es el predictor más utilizado, y la varianza condicionada mide el error que podemos cometer con este predictor.

Las distribuciones multivariantes más importantes son la multinomial, que generaliza la binomial, y la normal multivariante. Esta última distribución tiene la importante propiedad de que las relaciones existentes entre variables conjuntamente normales son siempre lineales.

Podemos generar valores al azar de una variable normal bivariante mediante el método de Montecarlo; primero generamos los valores de la pri-

mera componente a partir de su distribución marginal, como vimos en el capítulo 5, y después podemos generar un valor de la segunda utilizando la distribución condicionada de la segunda variable dado el valor de la primera.

6.10 Lecturas recomendadas

Casi todos los libros de cálculo de probabilidades incluidos en las referencias incluyen las distribuciones multivariantes. El material aquí presentado puede ampliarse en libros específicos de análisis multivariante, como Cuadras (1996), Johnson y Wichern (2007) y Peña (2002).

Apéndice 6A: El concepto de distancia y sus aplicaciones

Distancia euclídea

En geometría la distancia entre dos puntos es la longitud del segmento que los une. Dados los puntos **X**, de coordenadas (x_1, x_2, x_3), e $\mathbf{Y} = (y_1, y_2, y_3)$, su distancia se calcula por:

$$d = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2}$$

Esta noción puede extenderse a cualquier dimensión, y por analogía definimos la distancia entre dos puntos $(x_1, ..., x_n)$, $(y_1, ..., y_n)$ por la expresión:

$$d = \left(\sum_{i=1}^{n} [x_i - y_i]^2 \right)^{1/2}$$

que llamaremos distancia euclídea. Con esta distancia, los puntos equidistantes de un punto fijo, **X**, se encuentran en esferas con centro en **X**.

Distancia euclídea y análisis de datos

La distancia euclídea aparece de forma natural en el análisis de datos. Un conjunto de n datos $(x_1, ..., x_n)$ puede representarse como un punto en el espacio n-dimensional. En dicho espacio la constante a se representará como un vector con componentes iguales $(a, ..., a)$. La constante más próxima a un punto $(x_1, ..., x_n)$ será el valor que minimice

$$\Sigma (x_i - a)^2$$

es decir, su media aritmética.

La desviación típica es la distancia promedio entre los datos y su constante más próxima, la media aritmética. La covarianza entre **X** e **Y** es el cuadrado de la distancia euclídea entre los vectores $(\mathbf{X} - \mathbf{1}\bar{x})$ y $(\mathbf{Y} - \mathbf{1}\bar{x})$, que representan las desviaciones a la media.

Distancia entre variables multidimensionales

Si tenemos dos vectores $\mathbf{X}_1$, $\mathbf{X}_2$ que representan mediciones de distintas variables en dos individuos, la distancia euclídea es una medida poco adecuada cuando los componentes de estos vectores tienen distintas unidades: no es razonable sumar medidas en metros con otras en ptas. o en grados centígrados. Se utiliza entonces la *distancia estandarizada*, donde las diferencias entre las medidas se dividen por la desviación típica (para hacerlas adimensionales). Definimos la distancia estandarizada entre los vectores $(x_{11}, ..., x_{n1})$, $(x_{12}, ..., x_{n2})$, donde la variable x_{ij} tiene varianza σ_i, por:

$$D_s(\mathbf{X}_1, \mathbf{X}_2) = \left[\Sigma \left(\frac{x_{i1} - x_{i2}}{\sigma_i} \right)^2 \right]^{1/2}$$

Nótese que la distancia euclídea es un caso particular de ésta con $\sigma_i = \sigma = 1$.

Distancia de Mahalanobis

La distancia estandarizada no tiene en cuenta la posible dependencia entre las variables. Intuitivamente, si dos variables están muy relacionadas y en dos individuos toman valores muy distintos estos individuos deben considerarse más separados que si esa distancia se hubiese observado entre variables independientes. El cuadrado de la *distancia estandarizada* puede escribirse:

$$D_s^2(\mathbf{X}_1, \mathbf{X}_2) = (\mathbf{X}_1 - \mathbf{X}_2)' \mathbf{D}^{-1} (\mathbf{X}_1 - \mathbf{X}_2)$$

donde **D** es una matriz diagonal cuyos términos son las varianzas de las variables. Si en lugar de **D** utilizamos **M**, la matriz de varianzas y covarianzas entre las variables, obtenemos la *distancia de Mahalanobis*, definida por:

$$D_M^2(\mathbf{X}_1, \mathbf{X}_2) = (\mathbf{X}_1 - \mathbf{X}_2)' \mathbf{M}^{-1} (\mathbf{X}_1 - \mathbf{X}_2)$$

Las distancias estandarizada y euclídea son casos particulares poniendo $\mathbf{M} = \mathbf{D}$ o $\mathbf{M} = \mathbf{I}$. Esta distancia aparece naturalmente en estadística por su estrecha relación con la distribución normal. En efecto, el exponente de la función de densidad normal multivariante es:

$$(\mathbf{X} - \underline{\mu})' \mathbf{M}^{-1} (\mathbf{X} - \underline{\mu})$$

y representa la distancia de Mahalanobis entre cada punto y la media. Las curvas de nivel de esta distancia vendrán definidas por el conjunto de puntos:

$$cte = (\mathbf{X} - \underline{\mu})' \mathbf{M}^{-1} (\mathbf{X} - \underline{\mu})$$

y serán elipses con centro $\underline{\mu}$. Para la distancia euclidea $\mathbf{M}^{-1} = \mathbf{I}$, y las curvas de nivel son circunferencias.

Para aclarar este concepto, supongamos que tratamos de medir la distancia entre el aspecto físico de un grupo de personas y que tomamos la estatura (x) y el peso (y) para caracterizar a cada individuo. La medida de distancia estandarizada es:

$$\left(\frac{x_1 - x_2}{\sigma_1}\right)^2 + \left(\frac{y_1 - y_2}{\sigma_2}\right)^2$$

donde σ_1 y σ_2 son las desviaciones típicas de las variables x e y. Un problema de esta distancia es que no tiene en cuenta la dependencia entre ambas variables. Por ejemplo, tomando como referencia el individuo A (175 cm; 70 kg) y suponiendo $\sigma_1 = 5$ cm; $\sigma_2 = 5$ kg, los individuos B (185 cm; 80 kg) y C (165 cm; 80 kg) están a la misma distancia (8 unidades). Esto no es razonable, ya que el primer individuo es más alto, pero con proporciones similares, mientras que el segundo es más bajo, y mucho más gordo. Una medida mejor es tomar la distancia de Mahalanobis que en este caso se convierte en:

$$\left(\frac{1}{1-\rho^2}\right)\left[\left(\frac{x_1 - x_2}{\sigma_1}\right)^2 + \left(\frac{y_1 - y_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x_1 - x_2}{\sigma_1}\right)\left(\frac{y_1 - y_2}{\sigma_2}\right)\right]$$

Por tanto, si la relación entre la estatura y el peso es positiva, al movernos aumentando ambas, la distancia disminuye relativamente, mientras que al movernos en direcciones opuestas aumenta. Por ejemplo, con $\rho = 0{,}8$:

$$d(AB) = \frac{1}{1 - 0{,}8^2}\left[\left(\frac{10}{5}\right)^2 + \left(\frac{10}{5}\right)^2 - 2 \cdot (0{,}8)\left(\frac{10}{5}\right)\left(\frac{10}{5}\right)\right] = 4{,}4$$

$$d(AC) = \frac{1}{1 - 0{,}8^2}\left[\left(\frac{10}{5}\right)^2 + \left(\frac{10}{5}\right)^2 - 2 \cdot (0{,}8)\left(-\frac{10}{5}\right)\left(-\frac{10}{5}\right)\right] = 40$$

Indicando que la forma del individuo *A* está más próxima al *B* que al *C*, lo que concuerda con nuestra intuición.

Tercera parte

Inferencia

7. Estimación puntual

William Saely Gosset (Student)
(1876-1937)
Científico británico. Sus experimentos para mejorar la cerveza Guinness, compañía irlandesa para la que trabajó toda su vida, le llevaron a descubrir el estadístico *t* que lleva su nombre. Publicó su trabajo bajo el pseudónimo de Student ya que Guinness no permitía a sus empleados difundir el resultado de sus investigaciones.

7.1 Introducción a la inferencia estadística

La construcción de modelos probabilísticos presentada en los capítulos 4, 5 y 6 es un caso típico de razonamiento deductivo: se establecen hipótesis respecto al mecanismo generador de los datos y con ellas se deducen las probabilidades de los valores posibles. La inferencia estadística realiza el proceso inverso: dadas las frecuencias observadas de una variable, inferir el modelo probabilístico que ha generado los datos.

Los procedimientos de inferencia estadística pueden clasificarse por el objetivo del estudio, por el método utilizado y por la información considerada.

a) *Respecto al objetivo del estudio: muestreo frente a diseño*.
Cuando el objetivo es *describir* una variable o las relaciones entre un conjunto de variables, se utilizan técnicas de muestreo, que con-

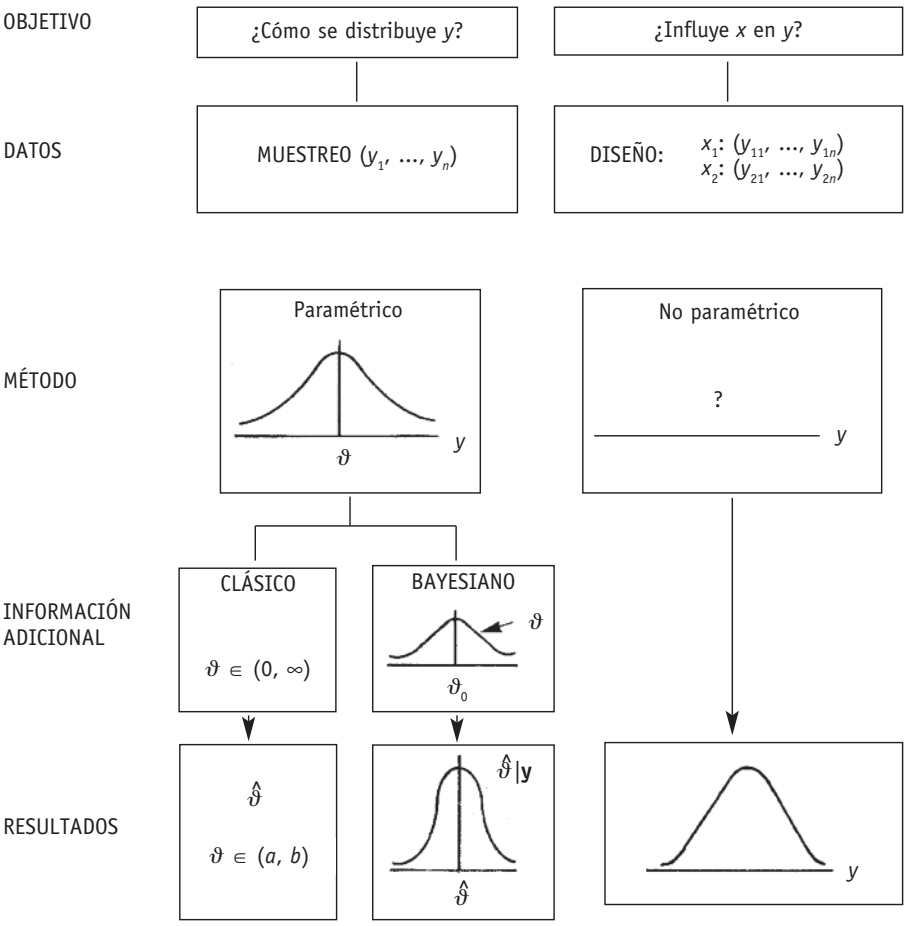
sisten en observar una muestra representativa de la población o poblaciones de interés. Cuando el objetivo es *contrastar* relaciones entre las variables y *predecir* sus valores futuros se utilizan técnicas de diseño experimental, que consisten en fijar los valores de ciertas variables y medir la respuesta que inducen en otras. En este primer tomo nos centraremos principalmente en métodos de muestreo; los métodos de diseño se abordarán con detalle en el segundo.

b) *Respecto al método utilizado: métodos paramétricos frente a no paramétricos*.
Los métodos paramétricos suponen que los datos provienen de una distribución que puede caracterizarse por un pequeño número de parámetros que se estiman a partir de los datos. Para ello suponen la forma de la distribución conocida (normal, Poisson, etc.) y deducen procedimientos óptimos para estimar sus parámetros.
Los métodos no paramétricos suponen únicamente aspectos muy generales de la distribución (que es continua, simétrica, etc.) y tratan de estimar su forma o contrastar su estructura. Dentro del enfoque paramétrico estos métodos se utilizan para contrastar hipótesis sobre la forma de la distribución.
En los capítulos 7 al 10 estudiaremos principalmente los métodos paramétricos de inferencia. Los métodos no paramétricos para contrastar la forma de la distribución y otras hipótesis se presentan en el capítulo 12.

c) *Respecto a la información considerada: enfoque clásico frente a bayesiano*.
El enfoque clásico supone que los parámetros son cantidades fijas desconocidas sobre los que no se dispone de información inicial relevante. Por tanto, la inferencia utiliza únicamente la información de los datos maestrales. El enfoque bayesiano considera los parámetros del modelo como variables aleatorias y permite introducir información inicial sobre sus valores mediante una distribución de probabilidad que se denomina distribución a priori. La diferencia práctica entre ambos procedimientos cuando disponemos de muestras grandes es muy escasa, ya que entonces la información de la muestra será siempre la determinante. En pequeñas muestras, sin embargo, pueden conducir a resultados distintos. El enfoque clásico se presenta en los capítulos 7 y 8, y el bayesiano, en el 9.

La figura 7.1 resume estas clasificaciones. Se estudia la distribución de una variable tomando una muestra (métodos de muestreo); se comprueba si x influye en y decidiendo unos valores de x y observando el comportamiento de y al cambiar x (métodos de diseño). En ambos casos es posible utilizar un método paramétrico o no paramétrico. En el primero se supone la forma de la distribución y se estima ϑ. En el segundo se estima directa-

Figura 7.1 Clasificación de los procedimientos de inferencia



mente la forma a partir de los datos (suavizando el histograma). Dentro del modelo paramétrico, si existe información relevante inicial sobre ϑ, podemos incluirla utilizando un enfoque bayesiano. Finalmente, el resultado del análisis será: (1) en el método clásico un estimador puntual de ϑ, $\hat{\vartheta}$, y un intervalo de valores posibles que indica la incertidumbre existente; (2) en el método bayesiano una distribución de probabilidad sobre ϑ; (3) en el enfoque no paramétrico una distribución estimada sobre y.

El método más común de inferencia es seleccionar la forma de la distribución inicial a la vista de los datos y luego aplicar un enfoque paramétrico (clásico o bayesiano) para estimar sus parámetros eficientemente.

7.2 Métodos de muestreo

7.2.1 Muestra y población

Llamaremos población a un conjunto homogéneo de elementos en los que se estudia una caracteristica dada. Frecuentemente no es posible estudiar todos ellos, ya que:

1) El estudio puede implicar la destrucción del elemento, como es el caso de ensayos destructivos: por ejemplo, estudiar la vida media de una partida de bombillas o la tensión de rotura de cables.
2) Los elementos pueden existir conceptualmente, pero no en la realidad. Por ejemplo, la población de piezas defectuosas que producirá una máquina.
3) Puede ser inviable económicamente estudiar toda la población.
4) El estudio llevaría tanto tiempo que sería impracticable, e incluso las propiedades de la población habrían variado con el tiempo.

En estas ocasiones en lugar de hacer un *censo* (un estudio exhaustivo de todos sus elementos) seleccionaremos un conjunto representativo de elementos que llamaremos *muestra*. Cuando la muestra está bien escogida podemos obtener una información similar a la del censo con mayor rapidez y menor coste. Esto justifica que, en la práctica, el análisis de poblaciones grandes se haga preferentemente mediante muestreo.

La clave de un procedimiento de muestreo es garantizar que la muestra sea representativa de la población. Por tanto, cualquier información respecto a las diferencias entre sus elementos debe tenerse en cuenta para seleccionar la muestra. Cuando no dispongamos de esta información y los elementos sean indistinguibles o intercambiables a priori y perfectamente homogéneos respecto a la variable que estudiamos, la muestra se selecciona con muestreo aleatorio simple, como describimos a continuación.

7.2.2 Muestreo aleatorio simple

Decimos que una muestra es aleatoria simple cuando:

1) Cada elemento de la población tiene la misma probabilidad de ser elegido.
2) Las observaciones se realizan con reemplazamiento, de manera que la población es idéntica en todas las extracciones.

La primera condición asegura la representatividad de la muestra: si el 20% de los elementos tiene la característica A y garantizamos con la forma

de seleccionar los elementos que todos tienen la misma probabilidad de aparecer, por término medio obtendremos un 20% de datos muestrales con la característica *A*.

La segunda condición se impone por simplicidad: si el tamaño de la población, *N*, es grande con relación al tamaño de la muestra *n*, es prácticamente indiferente realizar el muestreo con o sin reemplazamiento, pero el análisis resulta más simple cuando suponemos reemplazamiento. Si la fracción *n/N* es mayor que 0,1 (muestreamos más del 10% de la población), los métodos que presentamos son aproximados, y en el apéndice 7A se indican las correcciones pertinentes.

Para seleccionar una muestra por este método de una población finita se utilizan frecuentemente los números aleatorios de la forma siguiente: se numeran los elementos de la población de 1 a *N* y se toman números aleatorios de tantas cifras como tenga *N*. El valor del número aleatorio indicará el elemento a seleccionar.

En una muestra aleatoria simple cada observación tiene la distribución de probabilidad de la población. En efecto, cada observación es un valor al azar de la población y la probabilidad de que la observación sea menor que *A* coincidirá con la proporción de elementos de la población con valores menores que *A*. Sea $f(x)$ la distribución de la variable observada x y representemos la muestra por la variable n-dimensional

$$\mathbf{X}' = (x_1, ..., x_n)$$

donde x_i representa el valor de x en el elemento i-ésimo; entonces, llamando $f_1, \ldots, f_n$ a las funciones de densidad de estas variables, se verifica:

$$f_1 = f_2 = ... = f$$

Además, las observaciones son independientes y, por tanto, llamando f_c a la distribución conjunta de la muestra:

$$f_c(x_1, \ldots, x_n) = f_1(x_1) \ldots f_n(x_n) = f(x_1) \ldots f(x_n)$$

que es la condición matemática de muestra aleatoria simple.

7.2.3 Otros tipos de muestreo

Muestreo estratificado

El muestreo aleatorio simple debe utilizarse cuando los elementos de la población son homogéneos respecto a la característica a estudiar, es decir, a priori no conocemos qué elementos de la población tendrán valores altos de

ella. Cuando dispongamos de información sobre la población conviene tenerla en cuenta al seleccionar la muestra. Un ejemplo clásico son las encuestas de opinión, donde los elementos (personas) son heterogéneos en razón a su sexo, edad, profesión, etc. Interesa en estos casos que la muestra tenga una composición análoga a la población, lo que se consigue mediante una *muestra estratificada*.

Se denomina *muestreo estratificado* aquel en que los elementos de la población se dividen en clases o estratos. La muestra se toma asignando un número o *cuota* de miembros a cada estrato y escogiendo los elementos por muestreo aleatorio simple dentro del estrato.

En concreto, si existen k estratos de tamaños N_1, ..., N_k y tales que

$$N = N_1 + ... + N_k$$

tomaremos una muestra que garantice una presencia adecuada de cada estrato. Existen dos criterios básicos para dividir el tamaño total de la muestra (n) entre los estratos (n_i):

1) Proporcionalmente al tamaño relativo del estrato en la población (por ejemplo: si en la población hay 55% mujeres y 45% hombres, mantendremos esta proporción en la muestra). En general, $n_i = n \cdot (N_i/N)$.
2) Proporcionalmente a la variabilidad del estrato. Si conocemos la varianza de la característica a estudiar en cada estrato, tomaremos el tamaño muestral en cada uno proporcional a su variabilidad, de manera que los estratos más variables estén más representados. En concreto, si llamamos σ_i a la desviación tipica en el estrato i, se tomará:

$$n_i = n \cdot \frac{\sigma_i N_i}{\sum_{j=1}^{k} \sigma_j N_j}$$

que se reduce a la fórmula anterior si la variabilidad es aproximadamente constante.

Muestreo por conglomerados

Existen situaciones donde ni el muestreo aleatorio simple ni el estratificado son aplicables, ya que no disponemos de una lista con el número de elementos de la población ni de los posibles estratos. En estos casos típicamente los elementos de la población se encuentran de manera natural

agrupados en *conglomerados*, cuyo número sí se conoce. Por ejemplo, la población se distribuye en provincias, los habitantes de una ciudad en barrios, etc. Si podemos suponer que cada uno de estos conglomerados es una muestra representativa de la población total respecto a la variable que se estudia, podemos seleccionar algunos de estos conglomerados al azar y, dentro de ellos, analizar todos sus elementos o una muestra aleatoria simple. Este método se conoce como *muestreo por conglomerados* y tiene la ventaja de simplificar la recogida de la información muestral. El inconveniente obvio es que si los conglomerados son heterogéneos entre sí, como sólo se analizan algunos de ellos la muestra final puede no ser representativa de la población.

Por ejemplo, se desea tomar una muestra de la población española para estudiar la proporción de personas que están de acuerdo con las relaciones prematrimoniales. Si suponemos que la edad y el sexo pueden influir en la opinión, deberíamos tomar una muestra donde estas características sean las mismas que en la población base, lo que implica una muestra estratificada. Por otro lado, si suponemos que las provincias son homogéneas respecto a la opinión, podemos ahorrar muchos costes seleccionando al azar cuatro provincias y dentro de cada una de ellas una muestra aleatoria o, mejor, estratificada. Este procedimiento tiene el inconveniente obvio de que si las provincias no son homogéneas respecto a la opinión (por ejemplo las provincias más ricas tienen opinión distinta que las más pobres), tendremos sesgos (que evitaremos estratificando las provincias por riqueza).

En resumen, las ideas de *estratificación* y de *conglomerado* son opuestas: la estratificación funciona tanto mejor cuanto mayores sean las diferencias entre los estratos y más homogéneos sean éstos internamente; los conglomerados funcionan si hay muy pocas diferencias entre ellos y son muy heterogéneos internamente (incluyen toda la variabilidad de la población dentro de cada uno).

Muestreo sistemático

Cuando los elementos de la población están ordenados en listas, se utiliza el *muestreo sistemático*. Supongamos que la población tiene tamaño N y se desea una muestra de tamaño n. Sea k el entero más próximo a N/n. La muestra sistemática se toma eligiendo al azar (con números aleatorios) un elemento entre los primeros k. Sea n_1 el orden del elegido. Tomaremos a continuación los elementos $n_1 + k$; $n_1 + 2k$, etc., a intervalos fijos de k hasta completar la muestra. Si el orden de los elementos en la lista es al azar, este procedimiento es equivalente al muestreo aleatorio simple, aunque resulta más fácil de llevar a cabo sin errores. Si el orden de los elementos es tal que los individuos próximos tienden a ser más semejantes que los alejados,

el muestreo sistemático tiende a ser más preciso que el aleatorio simple, al cubrir más homogéneamente toda la población.

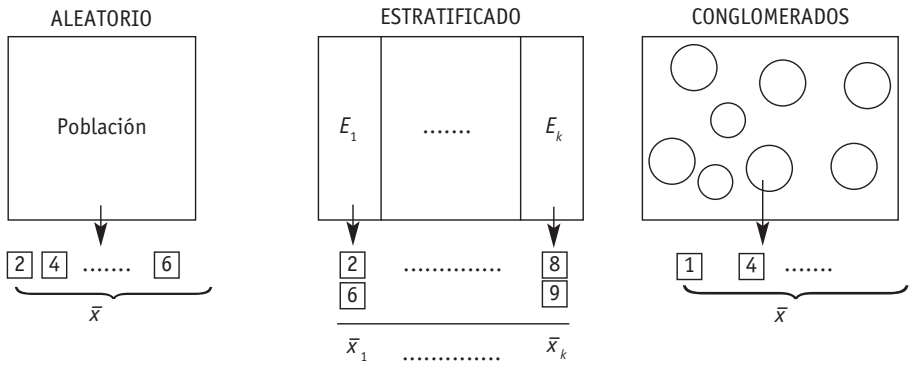
El muestreo sistemático puede utilizarse conjuntamente con el estratificado para seleccionar la muestra dentro de cada estrato.

Conclusión

La regla general que se aplica a todos los procedimientos de muestreo es que cualquier información previa debe utilizarse para subdividir la población y asegurar la mayor representatividad de la muestra. Una vez que disponemos de subpoblaciones homogéneas, la selección dentro de ellas debe realizarse por muestreo aleatorio simple.

En este libro supondremos siempre que la muestra proviene de un muestreo aleatorio simple. En el apéndice 7A se presenta brevemente el análisis en otros tipos de muestreo.

Figura 7.2 Diferencias entre el muestreo aleatorio, estratificado y por conglomerados



Ejercicios 7.1

7.1.1. Utilizando la tabla de números aleatorios del apéndice (o los generados por un ordenador), genere 50 muestras de las distribuciones:

a) Uniforme entre 10 y 20.
b) Exponencial con $\lambda = 2$.
c) Poisson con $\lambda = 1$.

Calcule la media muestral en cada una de las 50 muestras y haga un histograma de estos 50 valores. Comente el resultado obtenido.

7.1.2. Genere 100 muestras de 12 números aleatorios (x_1, ..., x_{12}) y calcule en cada muestra $y = x_1 + ... + x_{12} - 6$. Estudie la distribución de y. Compare la media y desviación típica observada con la teórica.

7.1.3. La llegada de aviones a un aeropuerto sigue una distribución de Poisson con parámetro $\lambda = 2$ llegadas/5 minutos. Genere una muestra de 3 horas de funcionamiento del aeropuerto utilizando el método de Montecarlo.

7.1.4. Elija al azar una página de la guía de teléfonos y cuente la distribución de frecuencias de los cuatro dígitos finales. ¿Aparecen las 10 cifras aproximadamente con la misma frecuencia?

7.1.5. Se desea realizar una encuesta para conocer la opinión de los estudiantes de una facultad o escuela respecto a la enseñanza que reciben. Indicar cómo seleccionar una muestra representativa para dicho estudio.

7.1.6. Tome una muestra sistemática de vocablos del diccionario de la Real Academia y cuente el número de palabras utilizadas para definirlos. Estime el número medio de palabras. Compare con otros diccionarios de español y de otros idiomas.

7.1.7. Indique un procedimiento para tomar una muestra de jóvenes entre 18 y 25 años de la población española para conocer su gasto en ocio.

7.3 La estimación puntual

7.3.1 Fundamentos

Supondremos en adelante en este capítulo que se observa una muestra aleatoria simple de una variable aleatoria x, que sigue una distribución conocida (normal, exponencial, Poisson, etc.), aunque con parámetros desconocidos. El problema que estudiaremos es cómo estimar estos parámetros a partir de los datos muestrales.

Supondremos que carecemos de información inicial respecto a los valores del parámetro ϑ. Cuando exista evidencia de que determinados valores del parámetro son mucho más probables que otros, utilizaremos el enfoque bayesiano, que se presenta en el capítulo 9.

El enfoque paramétrico supone que la forma del modelo es conocida. En la práctica, el tipo de variable a estudiar sugerirá una clase de modelos posibles, de la que seleccionaremos alguno a partir de la información previa disponible y del análisis de los datos muestrales. Vamos a comentar este aspecto más detalladamente.

7.3.2 La identificación del modelo

La primera operación a realizar con la muestra es un análisis descriptivo del tipo estudiado en el capítulo 2. Según la naturaleza de los datos construiremos un histograma, un diagrama de tallo y hojas o un diagrama de barras. Cuando la muestra sea grande (al menos 30 elementos), estas representaciones pueden ayudarnos a juzgar a priori si el modelo que estamos suponiendo es consistente con la muestra: si se ha supuesto normalidad, la muestra no debe reflejar claramente una distribución asimétrica, o valores separados de la media más de cuatro desviaciones típicas.

Con muestras pequeñas los gráficos anteriores son difíciles de interpretar. Por ello, se han diseñado gráficos en los que los puntos se sitúen en línea recta si el modelo supuesto es cierto. Vamos a presentar dos ejemplos de estos gráficos.

Gráfico para datos de Poisson

Si los datos siguen una distribución de Poisson, el valor esperado de las frecuencias observadas es:

$$E[f_{ob}(x)] = nP(x) = n\,\frac{e^{-\lambda}\lambda^{x}}{x!}$$

donde n es el tamaño muestral. Tomando logaritmos neperianos:

$$\ln E[f_{ob}(x)] = \ln n - \lambda + x \ln \lambda - \ln x!$$

Por tanto, si dibujamos $\ln f_{ob}(x) + \ln x!$ con respecto a x y los datos siguen una distribución de Poisson, la ecuación resultante será aproximadamente una recta, con pendiente $\ln \lambda$ y ordenada en el origen $\ln n - \lambda$. Una ventaja de este gráfico es que puede aplicarse aunque se desconozca la frecuencia de alguna clase.

Ejemplo 7.1

De Solla Price ha estudiado la distribución del número de descubrimientos científicos que han sido «redescubiertos» de forma independiente por otro autor, obteniendo los datos:

x Número de redescubrimientos	Frecuencia	$\ln f_{ob}(x) + \ln x!$
0	desconocida	—
1	no hay datos	—
2	179	5,88
3	51	5,72
4	17	6,01
5	6	6,58
mayor de 6	8	—

Llevando los cuatro puntos a un gráfico se obtiene aproximadamente una línea recta, lo que sugiere que la distribución de Poisson puede aceptarse como modelo de estos datos.

Figura 7.3 Gráfico de Poisson para los datos del ejemplo 7.1

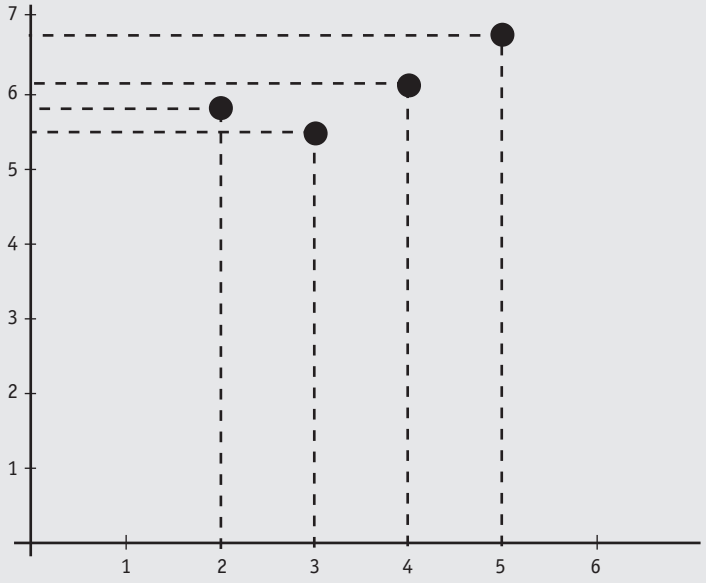
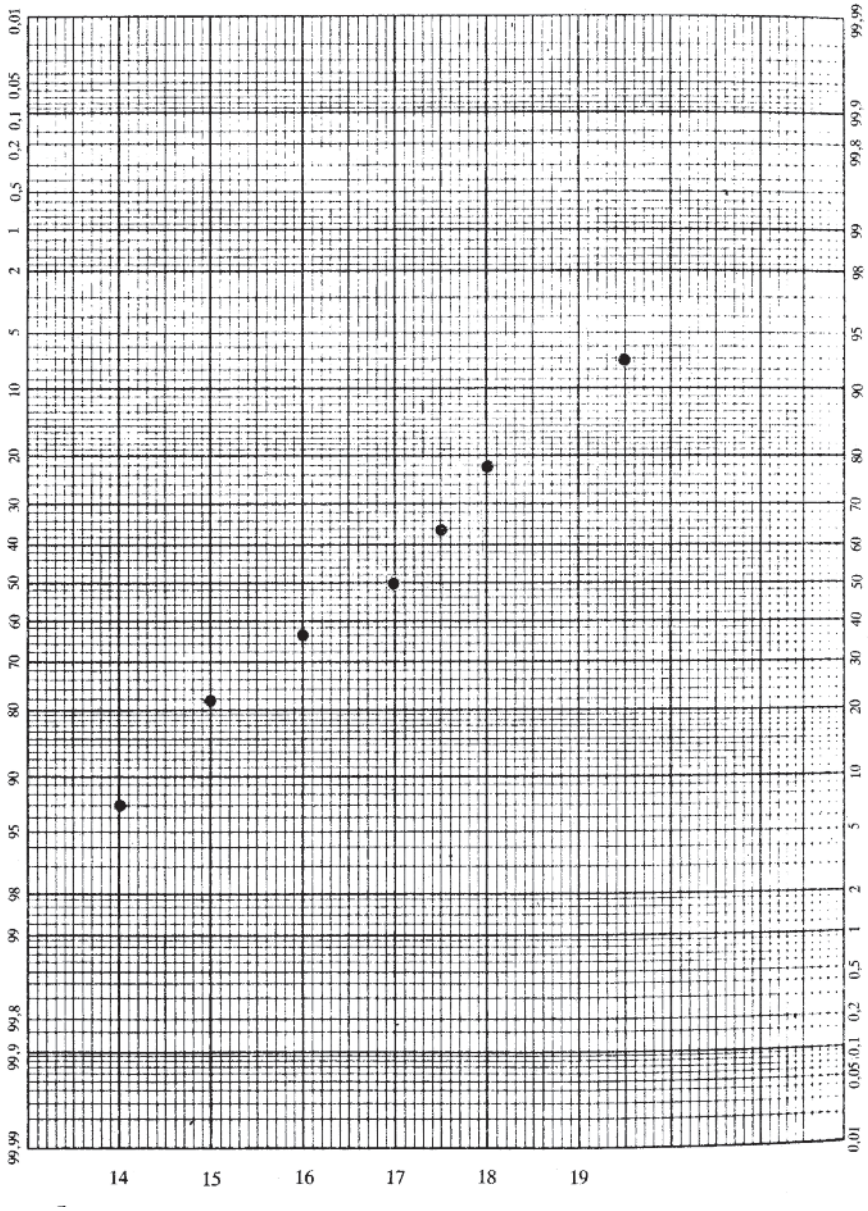


Gráfico para datos normales

El gráfico básico para datos normales utiliza el papel probabilístico normal. Se comienza construyendo la función de distribución empírica muestral, $F_n(x)$, definida por:

$$F_n(x_0) = fr(x \leq x_0)$$

Figura 7.4 Representación de la muestra 14; 17; 16; 15; 18; 19,5; 17,5; en papel probabilístico normal



En el caso en que no haya datos repetidos $F_n(x)$ toma los valores *1/n*; *2/n*; ..., 1. El papel probabilístico normal está construido de manera que el gráfico de x frente a $F_n(x)$ sea, si los datos son normales, una línea recta (véase la figura 7.4).

Cuando el tamaño muestral no es muy grande, el valor máximo observado en la muestra corresponde al valor 1 de F_n, lo que distorsiona la representación de los extremos. Para evitar este problema se recomienda:

a) Sea $x_{(1)} \leq x_{(2)} \ldots \leq x_{(i)} \ldots \leq x_{(n)}$ la muestra ordenada.
b) Dibujar $x_{(i)}$ en abscisas frente a $(i - 0{,}5)/n$ en ordenadas.

Si los puntos así dibujados se separan mucho de una recta, debe concluirse que la muestra no proviene de una distribución normal. Cuando los puntos centrales aparecen alineados pero no los extremos, hay que investigar si existen errores de datos u observaciones atípicas. Este punto se comentó en el capítulo 2 y volveremos sobre ello en el capítulo 12.

7.3.3 El método de los momentos

El primer método utilizado para obtener un *estimador* de un parámetro, es decir, un valor obtenido a partir de los datos muestrales, es el método de los momentos formalizado por K. Pearson a finales del siglo XIX. La idea es simple: tomar como estimador de la varianza de la población la varianza de la muestra; de la media de la población la media muestral, y así sucesivamente.

En general, si se trata de estimar un vector de parámetros $\underline{\vartheta} = (\vartheta_1, \ldots, \vartheta_k)$ cuyos componentes pueden expresarse en función de k momentos de la población, $m_1, \ldots m_k$, donde:

$$\begin{aligned} \vartheta_1 &= g_1(m_1, \ldots, m_k) \\ &\vdots \\ \vartheta_k &= g_k(m_1, \ldots, m_k) \end{aligned}$$

calcularemos los correspondientes momentos muestrales, $\hat{m}_1, \ldots, \hat{m}_k$, y los sustituiremos en el sistema de ecuaciones, para obtener los estimadores $\hat{\vartheta}_1, \ldots, \hat{\vartheta}_k$.

Para juzgar la bondad de los estimadores obtenidos por este procedimiento necesitamos establecer las propiedades deseables de los estimadores. Éste es el objeto de las secciones siguientes.

Ejemplo 7.2

Dada la muestra aleatoria (8; 6,5; 4; 7) de una χ^2, estimar sus grados de libertad por el método de los momentos.

Como $E[\chi^2] = n$, calculando la media $\hat{n} = \dfrac{8 + 6{,}5 + 4 + 7}{4} = 6{,}375$

Luego el estimador por momentos de n es 6.

Ejemplo 7.3

Dada la muestra (2, 4, 9, 1) de una distribución uniforme (0, b), estimar b.

$$\text{Como } E[x] = \frac{b}{2} \; ; \; \hat{b} = 2 \cdot \bar{x} = 2 \left(\frac{2 + 4 + 9 + 1}{4} \right) = 2 \cdot 4 = 8$$

El estimador obtenido en este caso no es muy razonable; es obvio que un estimador más preciso es el valor máximo observado en la muestra, 9.

7.4 La distribución de un estimador en el muestreo

7.4.1 Concepto

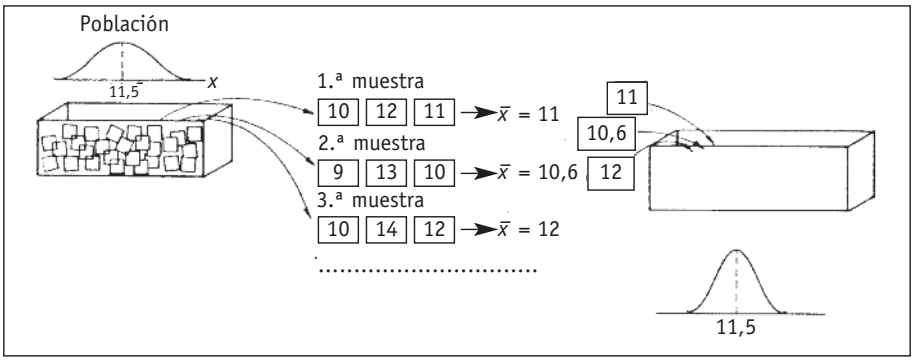
Al tratar de definir las propiedades de los estimadores nos encontramos con la dificultad de que el estimador es una variable aleatoria cuyo valor cambia de muestra en muestra. Consideremos una población de la que se toman muestras con reemplazamiento de tamaño n, calculando en cada muestra la media $\bar{x}$. En consecuencia, si tomamos k muestras obtendremos k valores, en general distintos, de medias muestrales $\bar{x}_1, ..., \bar{x}_k$. Si suponemos que k es muy grande —teóricamente infinito—, los valores $\bar{x}_i$ tendrán una distribución que llamaremos *distribución muestral de la media* en el muestreo.

La figura 7.5 ilustra esta situación. La población puede representarse por un cajón lleno de tarjetas, cada una de ellas con un valor de x. Para formar muestras de tamaño tres tomamos grupos de tres tarjetas al azar y calculamos las medias muestrales. Cuando esta operación se repite un número ilimitado de veces, la distribución de todas las medias muestrales así obtenidas la llamaremos distribución en el muestreo de la media muestral.

Observemos que la distribución en el muestreo de un estadístico depende de:

a) La población base.
b) El tamaño de muestra n.

Figura 7.5 La distribución muestral de la media



El cálculo matemático de la distribución de un estimador en el muestreo es, en general, complicado. Sin embargo, siempre podemos deducirla de manera aproximada, simulando con un ordenador el proceso de muestreo con el método de Montecarlo, como vimos en la sección 5.7. Este procedimiento se utiliza mucho en la práctica. En ciertos casos podemos acudir al teorema central del límite, que asegura que si el estimador es de la forma:

$$\hat{\vartheta} = a_1 x_1 + \ldots + a_n x_n$$

donde las a_i son constantes, tendrá, al aumentar n, una distribución asintóticamente normal.

En muchos casos las comparaciones entre estimadores no requieren en general conocer toda la distribución muestral, sino sólo sus principales momentos, que pueden calcularse directamente, como mostramos a continuación.

7.4.2 Distribución en el muestreo de una proporción

Supongamos una población donde observamos la presencia o no de un atributo. Sea p la proporción desconocida de elementos con dicho atributo. La distribución en el muestreo del estimador $\hat{p}$, proporción observada en la muestra, se obtiene inmediatamente de la distribución binomial. En efecto:

$$P(\hat{p} = r/n) = P_B(r) = \binom{n}{r} p^r (1-p)^{n-r} \qquad r = 0, 1, \ldots, n$$

Es decir, la probabilidad de que la proporción en la muestra sea r/n es igual a la probabilidad de obtener r elementos con esta característica en una

muestra de tamaño n, que es la distribución binomial. Por tanto, las propiedades de la distribución en el muestro del estimador $\hat{p}$ serán:

$$P[\hat{p}] = E[r/n] = \frac{np}{n} = p \tag{7.1}$$

$$Var[\hat{p}] = \frac{1}{n^2} 2\ Var[r] = \frac{pq}{n} \tag{7.2}$$

Además, cuando n sea grande, la distribución en el muestreo de $\hat{p}$ será aproximadamente normal con media y varianza dados por (7.1) y (7.2). Éste es un caso particular de la distribución muestral de una media, ya que $\hat{p}$ se calcula por:

$$\hat{p} = \frac{x_1 + ... + x_n}{n} \tag{7.3}$$

donde cada x_i toma el valor 1 si el elemento tiene el atributo estudiado y 0 en otro caso. Por tanto, $\hat{p}$ es la media muestral de las variables de Bernoulli, x_i.

7.4.3 Distribución muestral de la media

Vamos a calcular la media y varianza de la distribución muestral de la media en el caso general en que x es una variable aleatoria cualquiera con media μ y varianza σ^2. Entonces:

$$E[\bar{x}] = E\left[\frac{1}{n}\Sigma x_i\right] = \frac{1}{n}\Sigma E[x_i] = \frac{1}{n}\Sigma\mu = \mu$$

donde hemos utilizado que todas las variables x_i de una muestra aleatoria simple tienen la distribución de la población. La varianza de $\bar{x}$ será, utilizando que la varianza de una suma de variables aleatorias independientes es la suma de las varianzas de los sumandos:

$$Var[\bar{x}] = \frac{1}{n^2}\Sigma Var[x_i] = \frac{\sigma^2}{n}$$

Por tanto, concluimos que al tomar una muestra de tamaño n de una variable con media μ, varianza σ^2 y distribución cualquiera, la distribución muestral de la media verifica:

$$\boxed{E[\bar{x}] = \mu \qquad Var[\bar{x}] = \sigma^2/n} \tag{7.4}$$

Observemos que los resultados (7.1) y (7.2) son casos particulares de (7.4) cuando las variables x son de Bernoulli, con media p y varianza $p \cdot q$.

La distribución exacta de $\bar{x}$ para pequeñas muestras depende de la población. Por ejemplo, si x es normal, la distribución de $\bar{x}$ lo será también por ser una combinación lineal de variables normales. Además, asintóticamente la distribución de $\bar{x}$ será normal, en virtud del teorema central del límite, sea cual sea la distribución de la población de partida (excluyendo casos patológicos) (véase la figura 7.6). En la práctica la aproximación normal se utiliza cuando $n \geq 30$.

La distribución de la media es siempre más simétrica que la distribución original de la variable. Puede comprobarse (véase ejercicio 7.2.8) que el coeficiente de asimetría de la distribución de la media muestral es igual al coeficiente de asimetría de la población dividido por $\sqrt{n}$.

7.4.4 Distribución muestral de la varianza. Caso general

Esperanza

La esperanza de la distribución muestral de la varianza de una variable aleatoria cualquiera con media μ y varianza σ^2 será:

$$E[s^2] = \frac{1}{n} \sum E[(x_i - \bar{x})^2] \tag{7.5}$$

como:

$$\sum (x_i - \bar{x})^2 = \sum (x_i - \mu + \mu - \bar{x})^2 = \sum (x_i - \mu)^2 + n(\mu - \bar{x})^2 + 2(\mu - \bar{x}) \sum (x_i - \mu)$$

resulta que:

$$\sum (x_i - \bar{x})^2 = \sum (x_i - \ \mu)^2 - n(\mu - \bar{x})^2$$

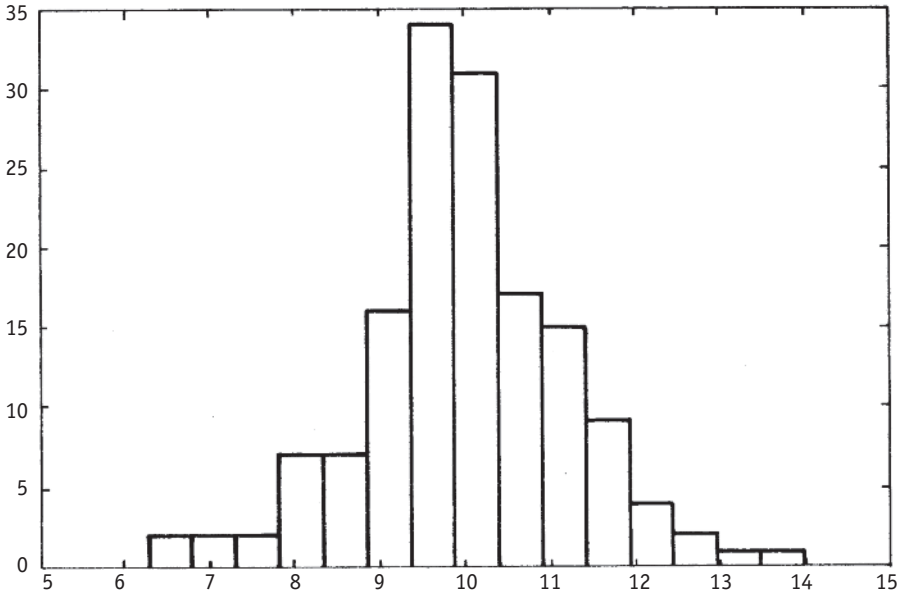
que escribiremos:

$$\boxed{\sum (x_i - \mu)^2 = \sum (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2} \tag{7.6}$$

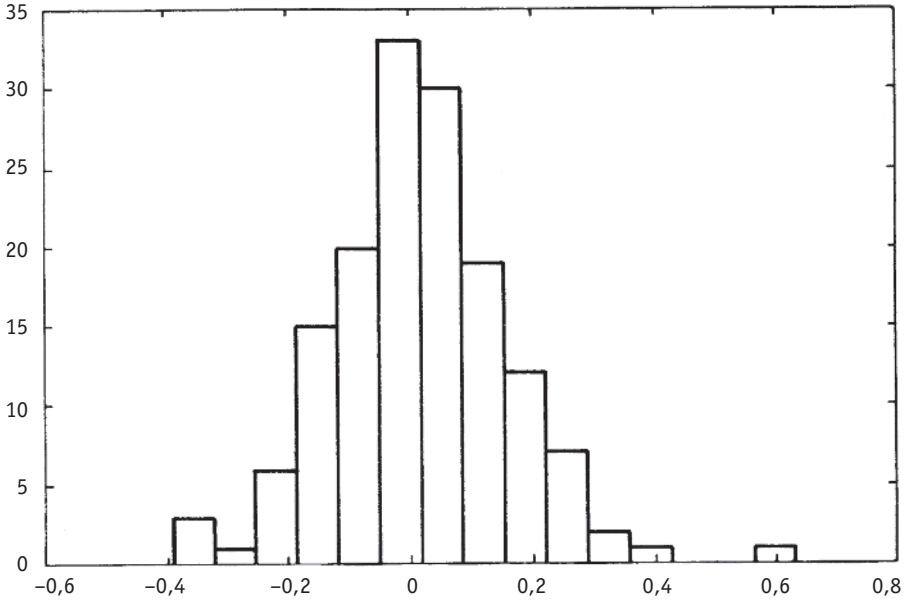
Este resultado tiene una importante interpretación: descompone la variabilidad de los datos respecto a su media verdadera como suma de la variabilidad respecto a la media muestral y la variabilidad entre la media muestral y la verdadera. Tomando esperanzas en (7.6)

$$n\sigma^2 = E[ns^2] + nE[(\bar{x} - \mu)^2]$$

Figura 7.6 Histograma de 150 muestras con $n = 50$ de la distribución en el muestreo de la media. (a) Población exponencial ($\lambda = 10$); (b) población normal $N(0, 1)$



(a)



(b)

como según (7.4)

$$E[(\bar{x} - \mu)^2] = Var(\bar{x}) = \frac{\sigma^2}{n}$$

resulta que

$$E[s^2] = \sigma^2 - \frac{\sigma^2}{n} = \sigma^2 \frac{n-1}{n} \tag{7.7}$$

En consecuencia, el valor medio de s^2 es menor que σ^2, aunque la diferencia tiende a cero al aumentar n.

La varianza muestral corregida

Si definimos la varianza muestral corregido por:

$$\hat{s}^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{n}{n-1} s^2 \tag{7.8}$$

se verifica, según (7.7), que:

$$E[\hat{s}^2] = \sigma^2 \tag{7.9}$$

El divisor, $n - 1$, se denomina *número de grados de libertad* y tiene en cuenta el número de términos desconocidos antes de tomar la muestra que incluimos en el cálculo del estimador. Para interpretar esta importante idea, llamaremos residuo a:

$$\text{Residuo} = e_i = x_i - \bar{x} \tag{7.10}$$

la diferencia entre el valor observado y el estimado. Entonces, la varianza muestral corregido se calcula:

$$\hat{s}^2 = \frac{\sum e_i^2}{n-1} \tag{7.11}$$

Cuando $n = 1$, $\bar{x} = x_1$ y antes de tomar la muestra podemos afirmar que $e_1 = 0$. No hay ningún grado de libertad. Si $n = 2$, tendremos que:

$$e_1 = x_1 - \bar{x} = x_1 - \left(\frac{x_1 + x_2}{2}\right) = x_1 - x_2 = -(x_2 + x_1) = -e_2$$

y hay solamente un grado de libertad: el valor de e_1 (*o* de e_2). Dado un residuo, el otro queda automáticamente fijado. En general, como para cualquier tamaño muestral

$$\Sigma(x_i - \bar{x}) = \Sigma e_i = 0 \tag{7.12}$$

antes de tomar la muestra sólo hay $n - 1$ residuos desconocidos porque el último siempre puede calcularse con (7.12). Diremos que disponemos de $n - 1$ grados de libertad para calcular los residuos y, por tanto, la desviación típica de los datos.

En resumen, la varianza muestral, s^2, tiende a subestimar, por término medio, la varianza de la población. Este resultado es debido a que en vez de calcular las desviaciones $\Sigma(x_i - \mu)^2$ al ser μ desconocida calculamos $\Sigma(x_i - \bar{x})^2$ que, según (7.6), será siempre menor. Para corregir por este efecto dividimos por $n - 1$, que es el número de grados de libertad de los residuos.

Distribución

La distribución de s^2 (o de $\hat{s}^2$) es típicamente asimétrica (véase la figura 7.7) y su forma depende de n, tamaño muestral, y de la población base. Según el teorema central del límite, tenderá asintóticamente a la normal, pero la convergencia es muy lenta y sólo se manifiesta para tamaños muestrales grandes. Sin embargo, su logaritmo tiene, en general, una distribución más simétrica.

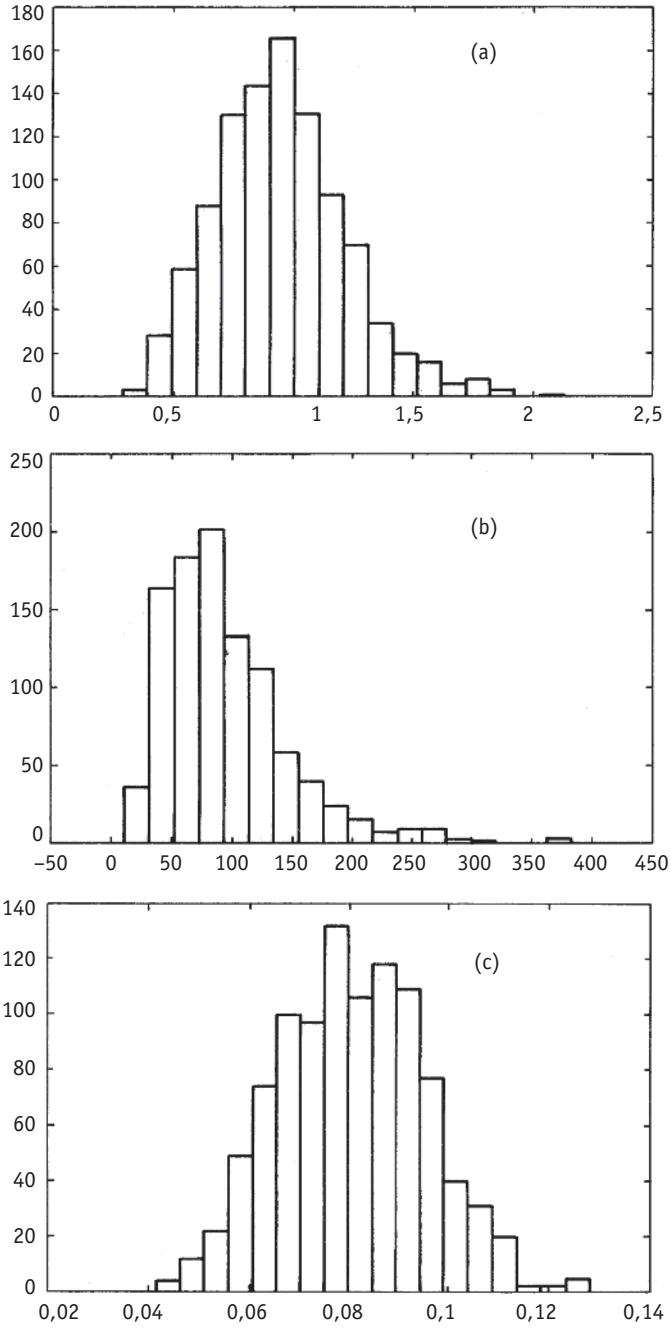
7.4.5 Distribución muestral de la varianza en poblaciones normales

Si la población base es normal, dividiendo por σ^2 en (7.6) obtenemos la expresión:

$$\Sigma\left(\frac{x_i - \mu}{\sigma}\right)^2 = \Sigma\left(\frac{x_i - \bar{x}}{\sigma}\right)^2 + \left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}\right)^2 \tag{7.13}$$

el primer miembro es la suma de cuadrados de n variables aleatorias $N(0, 1)$ independientes, y será por tanto una χ^2 con n grados de libertad. En el segundo miembro el último término es el cuadrado de otra variable $N(0, 1)$, según (7.4). El término $\Sigma(x_i - \bar{x})^2/\sigma^2$ es la suma de n variables $x_i - \bar{x}$ que están ligadas por la restricción (7.12), con lo que tendrá $n - 1$ grados de libertad. Puede demostrarse que este término puede escribirse

Figura 7.7 **Histograma de s^2 en 1.000 muestras de tamaño $n = 25$ extraídas de una población (a) $N(0, 1)$; (b) exponencial con $\lambda = 0{,}1$; (c) Uniforme (0, 1)**



como la suma de $n-1$ variables normales, independientes, y, por tanto, seguirá una distribución χ^2 con $n-1$ grados de libertad. Este resultado puede expresarse así:

$$\boxed{\frac{\hat{s}^2}{\sigma^2} \sim \frac{\chi^2_{n-1}}{n-1}} \tag{7.14}$$

Además se comprueba que las variables aleatorias $\hat{s}$ y $\bar{x}$, la media y la desviación típica muestrales, son independientes. Esta propiedad caracteriza a la distribución normal: en cualquier otra distribución estos estimadores son dependientes (dependencia positiva implica que si la media es, por azar, anormalmente alta en la muestra, también lo será la desviación típica y al revés). Escribiremos:

$$\boxed{\bar{x} \text{ y } \hat{s} \text{ son independientes}}$$

Por último, utilizando las propiedades de la χ^2 es inmediato comprobar que:

$$E[\hat{s}^2] = E\left[\frac{\sigma^2}{n-1}\chi^2_{n-1}\right] = \sigma^2 \tag{7.15}$$

y este estimador es siempre centrado. Por tanto, en promedio el cociente $\hat{s}^2/\sigma^2$ es igual a la unidad. Además:

$$Var[\hat{s}^2] = Var\left[\frac{\sigma^2}{n-1}\chi^2_{n-1}\right] = \frac{2\sigma^4}{(n-1)} \tag{7.16}$$

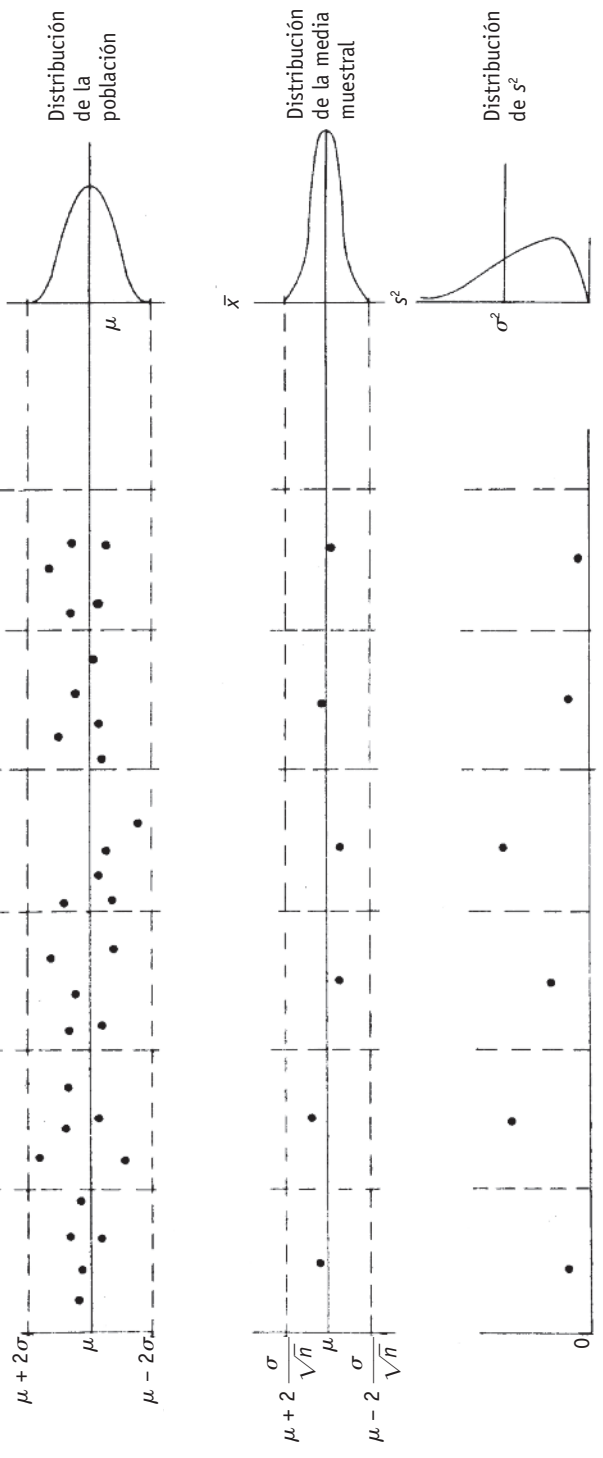
que puede también expresarse diciendo que el cociente $\hat{s}^2/\sigma^2$ tiene una desviación típica $\sqrt{2/(n-1)}$.

Distribución muestral de la desviación típica

Para poblaciones normales la desviación típica muestral sigue una distribución relacionada con la χ^2. Utilizando la relación aproximada entre los momentos de una variable y su transformada y el resultado (7.15):

$$E[\hat{s}] = \sigma + Var(\hat{s}^2) \cdot 1/2[-1/4(\sigma^2)^{-3/2}]$$

Figura 7.8 Distribución de la media y la varianza muestral en poblaciones normales (adaptada de Box, Hunter y Hunter, 1978)



y sustituyendo (7.16) en esta ecuación:

$$E[\hat{s}] = \left[1 - \frac{1}{4(n-1)}\right]\sigma \tag{7.17}$$

que indica que $\hat{s}$ subestima algo σ por término medio, aunque para tamaños muestrales medianos dicho efecto es despreciable. Aunque esta fórmula es sólo aproximada (véase el apéndice 7A para el resultado exacto), el error es pequeño. Por otro lado,

$$Var(\hat{s}) = Var(\hat{s}^2) \cdot \left[\frac{1}{2\sqrt{\sigma^2}}\right]^2 = \frac{\sigma^2}{2(n-1)} \tag{7.18}$$

resultado de nuevo aproximado para n pequeño pero bastante preciso en muestras grandes. El cuadro 7.1 resume estas propiedades.

Cuadro 7.1 Medias y varianzas muestrales de los estadísticos más frecuentes (los valores de $\bar{x}$ y $\hat{s}^2$ son exactos, los de $\hat{s}$ buenas aproximaciones)

Estadístico	Media	Varianza (general)	Varianza Poblaciones normales
$\bar{x}$	μ	σ^2/n	σ^2/n
$\hat{s}^2$	σ^2	$\sigma^4\left(\frac{2}{n-1} + \frac{(CA_p - 3)}{n}\right)$	$\frac{2\sigma^4}{n-1}$
$\hat{s}$	$\frac{4n-5}{4n-4}\sigma$	—	$\frac{\sigma^2}{2(n-1)}$

Ejercicios 7.2

7.2.1. Obtenga una muestra aleatoria siempre de tamaño 20 de sus compañeros de curso y estudie las variables siguientes:

a) Estatura.
b) Tiempo que invierten en desplazamiento.
c) Gastos semanales.
d) Número de días en cama por enfermedad el curso pasado.

Proponga un modelo para cada variable y estudie gráficamente la concordancia de la muestra con el modelo. (Intente transformaciones si lo considera conveniente.)

7.2.2. Utilizando las variables $y = x_1 + ... + x_{12} - 6$, donde x_1, ..., x_{12} son números aleatorios (véase el ejercicio 7.1.2), seleccione cinco muestras aleatorias simples de tamaño 20 de una población $N(10, 2)$. Dibuje cada muestra en papel probabilístico normal. Comente el resultado obtenido.

7.2.3. Los taxis en servicio de una ciudad están numerados del 1 al N. Se observa una muestra de 10 taxis y se apuntan sus números. Obtener un estimador de N por el método de los momentos.

7.2.4. Se ha analizado un conjunto de n microprocesadores y se encuentran x defectuosos. No se conoce n, pero sí la probabilidad de defecto p. Estimar n por el método de los momentos.

7.2.5. La vida de un mecanismo es una variable aleatoria con densidad $f(x) = (x/a^2)e^{-x^2/2a^2}$ para $x > 0$ (distribución de Weibull). Encontrar un estimador por momentos de a.

7.2.6. Estudie por el método de Montecarlo la distribución de la varianza muestral de una población normal. Compare la distribución con la de $\ln s^2$.

7.2.7. Obtener por el método de los momentos un estimador para el parámetro a en:
a) $f(x) = \frac{2}{a^2}(a - x)$ para $0 < x < a$.
b) $f(x) = ax_0^a/x^{a+1}$ $(x > x_0)$. (Ésta es la distribución de Pareto, utilizada en el análisis de la distribución de la renta.)

7.2.8. Demostrar que el coeficiente de asimetría de la distribución muestral de la media es igual al de la población dividido por la raíz del tamaño muestral.

7.5 Propiedades de los estimadores

7.5.1 Centrado o insesgado

Diremos que un estimador $\hat{\vartheta}$ es centrado o insesgado para ϑ, si para cualquier tamaño muestral,

$$\boxed{E[\hat{\vartheta}] = \vartheta}$$

Cuando el estimador no es centrado, se define:

$$\text{sesgo}\,(\hat{\vartheta}) = E(\hat{\vartheta}) - \vartheta$$

Pueden existir muchos estimadores centrados para un parámetro. Por ejemplo, para estimar μ en una distribución cualquiera, todos los estimadores del tipo

$$\hat{\mu} = a_1 x_1 + \ldots + a_n x_n \qquad \text{con} \qquad \sum a_i = 1$$

son centrados.

En la sección anterior hemos comprobado que $\bar{x}$ (y como caso particular $\hat{p}$) es siempre centrado para estimar μ y que s^2 *no* es centrado para estimar σ^2. Aunque es posible que el sesgo dependa del parámetro desconocido, ϑ, en general podemos conocer a priori si el estimador es centrado o no. Por otro lado es frecuente que el sesgo dependa del tamaño muestral, como hemos visto en el caso de la varianza.

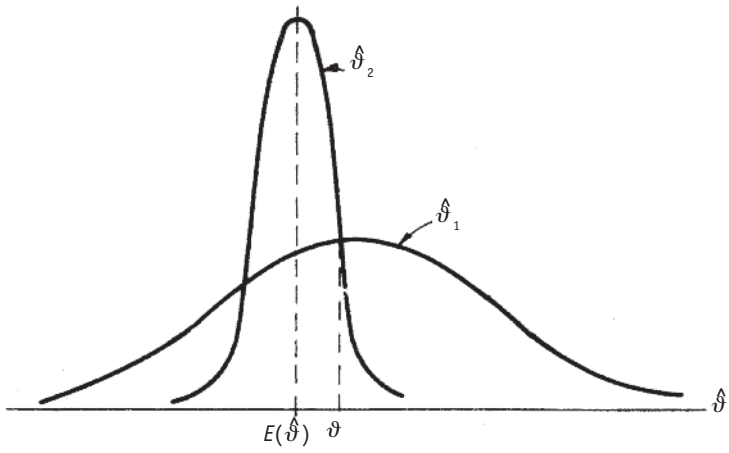
Una ventaja adicional de los estimadores centrados es que podemos combinarlos para obtener nuevos estimadores centrados: si tenemos dos muestras independientes y calculamos en cada una de ellas un estimador centrado $\hat{\vartheta}_i$ para el parámetro, cualquier estimador del tipo:

$$\hat{\vartheta}_T = a_1 \hat{\vartheta}_1 + a_2 \hat{\vartheta}_2 \qquad ; \qquad a_1 + a_2 = 1$$

será también centrado.

La propiedad de ser centrado no es por sí sola concluyente. Por ejemplo, la figura 7.9 muestra dos estimadores: el primero es centrado, pero con gran varianza, por lo que el segundo será preferido aunque sea sesgado.

Figura 7.9 Comparación de dos estimadores de un parámetro ϑ



7.5.2 Eficiencia o precisión

Llamaremos *eficiencia* o *precisión* de un estimador a la inversa de la varianza de su distribución muestral. Es decir:

$$\boxed{\text{precisión}\ (\hat{\vartheta}) = 1/Var(\hat{\vartheta})} \qquad (7.19)$$

Diremos que un estimador $\hat{\vartheta}_2$ es más eficiente o más preciso que otro $\hat{\vartheta}_1$ si para cualquier tamaño muestral (figura 7.9):

$$Var(\hat{\vartheta}_2) \leq Var(\hat{\vartheta}_1) \Leftrightarrow \text{efic}\ (\hat{\vartheta}_2) \geq \text{efic}\ (\hat{\vartheta}_1)$$

llamaremos *eficiencia relativa* de $\hat{\vartheta}_2$ respecto a $\hat{\vartheta}_1$ al cociente entre sus eficiencias:

$$ER\ (\hat{\vartheta}_2/\hat{\vartheta}_1) = \frac{\text{efic}\ (\hat{\vartheta}_2)}{\text{efic}\ (\hat{\vartheta}_1)} = \frac{Var\ (\hat{\vartheta}_1)}{Var\ (\hat{\vartheta}_2)}$$

La eficiencia es pues un concepto ligado a la varianza, y es especialmente relevante para comparar estimadores centrados, ya que, entre ellos, será preferido el más eficiente.

Por ejemplo, en poblaciones normales la mediana muestral es un estimador centrado de la media de la población, con varianza asintótica $(\pi\sigma^2)/2n$. La eficiencia relativa de la media muestral respecto a la mediana en muestras grandes es:

$$ER(\text{Media/Mediana}) = \frac{(\pi\sigma^2)/2n}{\sigma^2/n} = 1{,}57$$

La varianza de la mediana muestral es un 57% más alta que la de la media. Esto implica que la precisión de la media muestral con $n = 100$ es equivalente a la de la mediana muestral con $2n/\pi = 100$, es decir, con $n = 157$. En general, si la eficiencia de un estimador respecto a otro es 2, esto implica que necesitamos con el segundo un tamaño muestral doble para tener la misma precisión (varianza) que con el primero.

Combinación lineal de estimadores centrados

Si se toman distintas muestras independientes y se calcula en cada una un estimador centrado de un mismo parámetro, se presenta el problema de cómo combinar estos estimadores independientes para obtener el mejor estimador que sintetice toda la información disponible. Por ejemplo, se dispo-

ne de dos encuestas que proporcionan valores distintos de la proporción de votantes de un partido o usuarios de un producto; o los resultados de dos laboratorios que han obtenido estimaciones distintas al medir una misma magnitud.

Para simplificar, supongamos dos muestras independientes que dan lugar a los dos estimadores, $\hat{\vartheta}_1$, $\hat{\vartheta}_2$. Entonces cualquier estimador del tipo:

$$\hat{\vartheta}_T = a\hat{\vartheta}_1 + (1-a)\hat{\vartheta}_2$$

será centrado. Para determinar el de menor varianza (mayor precisión), como:

$$\text{Var}(\hat{\vartheta}_T) = a^2\,\text{Var}(\hat{\vartheta}_1) + (1-a)^2\,\text{Var}(\hat{\vartheta}_2)$$

Derivando respecto a para determinar el valor mínimo de esta varianza:

$$\frac{d\,\text{Var}(\hat{\vartheta}_T)}{da} = 0 = 2a\,\text{Var}(\hat{\vartheta}_1) - 2(1-a)\,\text{Var}(\hat{\vartheta}_2)$$

que resulta ser:

$$a = \frac{\text{Var}(\hat{\vartheta}_2)}{\text{Var}(\hat{\vartheta}_1) + \text{Var}(\hat{\vartheta}_2)} = \frac{\text{Prec}(\hat{\vartheta}_1)}{\text{Prec}(\hat{\vartheta}_1) + \text{Prec}(\hat{\vartheta}_2)}$$

Este resultado ilustra la siguiente conclusión general: *la combinación lineal más precisa de estimadores centrados independientes es la construida con ponderaciones directamente proporcionales a la precisión relativa de cada estimador*.

Por ejemplo, si disponemos de tres estimadores independientes $\hat{p}_1$, $\hat{p}_2$, $\hat{p}_3$, de un parámetro p, el peso de cada una será:

$$a_i = \frac{\text{Prec}(i)}{\sum \text{Prec}(j)} = \frac{n_i/pq}{n_1/pq + n_2/pq + n_3/pq} = \frac{n_i}{\sum n_j}$$

y el estimador final será:

$$\hat{P}_T = \frac{n_1\hat{p}_1 + n_2\hat{p}_2 + n_3\hat{p}_3}{n_1 + n_2 + n_3}$$

que equivale a contar el número total de elementos con la característica estudiada en las tres muestras y dividir por el número total de elementos estudiados.

Ejemplo 7.4

Para estimar las ventas medias diarias se han tomado muestras de tres meses distintos de 20, 22 y 18 días laborables respectivamente obteniendo ventas medias de 200, 180 y 210 y desviaciones típicas corregidas de 52, 46 y 38 respectivamente. Si suponemos que las ventas son estables (no hay tendencia creciente ni decreciente), homogéneas en los meses (no hay estacionalidad) y con la misma variabilidad promedio en todos los meses, estimar la media de ventas diaria y la desviación típica.

Las estimaciones 200, 180 y 210 tienen una desviación típica de $\sigma/\sqrt{20}$, $\sigma/\sqrt{22}$ y $\sigma/\sqrt{18}$ respectivamente. Suponiendo que σ es la misma en todos los meses, tendremos:

$$\bar{x} = \frac{20}{20+22+18}(200) + \frac{22}{20+22+18}(180) + \frac{18}{20+22+18}(210) = 195{,}66$$

Las varianzas muestrales corregidas son estimadores centrados con varianza en poblaciones normales aproximadamente proporcionales a los tamaños muestrales menos 1. Entonces:

$$\hat{\sigma}^2 = \frac{19}{19+21+17} \cdot 52^2 + \frac{21}{19+21+17} \cdot 46^2 + \frac{17}{19+21+17} \cdot 38^2 = 2.111{,}6$$

es un estimador de σ^2. Tomando como estimador σ su raíz, que es prácticamente centrado, obtenemos que $\hat{s} = 45{,}95$.

7.5.3 Error cuadrático medio

A veces se presenta el problema de elegir entre dos estimadores con propiedades contrapuestas: uno de ellos, $\hat{\vartheta}_1$, es centrado, mientras que el otro, $\hat{\vartheta}_2$, es sesgado, aunque con menor varianza. En estos casos, es razonable elegir aquel estimador con menor error promedio de predicción del parámetro. Por definición:

$$\boxed{ECM(\vartheta) = E\,[(\hat{\vartheta} - \vartheta)^2]} \qquad (7.20)$$

donde *ECM* significa error cuadrático medio, y el promedio se toma con respecto a la distribución en el muestreo del estimador $\hat{\vartheta}$. Se verifica:

$$E[(\hat{\vartheta} - \vartheta)^2] = E[(\hat{\vartheta} - E[\hat{\vartheta}] + E[\hat{\vartheta}] - \vartheta)^2] = (E[\hat{\vartheta}] - \vartheta)^2 + E(\hat{\vartheta} - E[\hat{\vartheta}])^2$$

ya que como $E(\hat{\vartheta}) - \vartheta$ es una constante, coincide con su esperanza y el doble producto se anula; por tanto:

$$\boxed{ECM(\hat{\vartheta}) = [\text{sesgo}\,(\hat{\vartheta})]^2 + \text{Var}(\hat{\vartheta})} \tag{7.21}$$

y para estimadores centrados el error cuadrático medio coincide con la varianza.

Aunque en general el error cuadrático medio depende de ϑ y del tamaño muestral, es frecuente al comparar estimadores que uno tenga menor error cuadrático medio para cualquier valor de ϑ y tamaño muestral. Entonces diremos que el estimador con mayor ECM es *inadmisible* con relación a este criterio.

Ejemplo 7.5

Comparar los estimadores s^2 y $\hat{s}^2$ desde el punto de vista de sus errores cuadráticos medios.

El sesgo de s^2 es:

$$\text{sesgo}(s^2) = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{\sigma^2}{n}$$

y será negativo ya que s^2 en promedio subestima σ^2. Su varianza es:

$$\text{Var}(s^2) = \text{Var}\left(\frac{n-1}{n}\hat{s}^2\right) = \frac{(n-1)^2}{n^2} \cdot \frac{2\sigma^4}{(n-1)} = \frac{2(n-1)\sigma^4}{n^2}$$

y el error cuadrático medio será:

$$ECM(s^2) = \frac{\sigma^4}{n^2} + \frac{2(n-1)}{n^2}\sigma^4 = \frac{\sigma^4}{n^2}2(n-1)$$

Como $\hat{s}^2$ es centrado, su *ECM* es directamente su varianza:

$$ECM(\hat{s}^2) = \frac{2}{(n-1)}\sigma^4$$

y como:

$$\frac{2}{n-1} > \frac{2}{n} > \frac{2}{n} - \frac{1}{n^2}$$

Por tanto, con este criterio, el estimador sin corregir es preferible. La diferencia entre ambos estimadores es pequeña cuando n es grande.

7.5.4 Consistencia

Cuando disponemos de muestras grandes y no sea posible —o sea difícil— la obtención de estimadores centrados con alta eficiencia, el requisito mínimo que se exige a un estimador es que sea *consistente*, entendiendo por ello que se aproxime, al crecer el tamaño muestral, al valor del parámetro. Intuitivamente, diremos que la secuencia de estimadores $\hat{\vartheta}_n$ es consistente si, al aumentar n:

$$E[\hat{\vartheta}_n] \to \vartheta$$

es decir, la esperanza del estimador es, asintóticamente, el valor del parámetro, y:

$$\text{Var}(\hat{\vartheta}_n) \to 0$$

que indica que la varianza tiende a cero con n.

Esta definición de consistencia —que estrictamente se denomina consistencia en media cuadrática— es más restrictiva de lo necesario para garantizar la aproximación de $\hat{\vartheta}$ hacia ϑ al aumentar n, pero es operativa y simple y la seguiremos en el resto del libro. Existen otras definiciones de consistencia que el lector interesado puede encontrar en cualquier libro de estadística matemática.

7.5.5 Robustez

Concepto

Una propiedad deseable de un buen estimador para un parámetro ϑ en el modelo $f(x)$ es continuar siendo razonablemente bueno como estimador de ϑ si el modelo experimenta una pequeña modificación. Cuando esto ocurre, diremos que el estimador es *robusto* para ϑ. En concreto, consideremos alteraciones del modelo $f(x)$ del tipo:

$$(1 - \alpha)f(x) + \alpha g(x)$$

donde α es un valor positivo pequeño (0,01 o 0,001). Intuitivamente esta ecuación expresa la función de densidad de una variable que se genera con alta probabilidad $(1 - \alpha)$ de la distribución supuesta, $f(x)$, y con pequeña probabilidad α de otra distribución arbitraria, $g(x)$.

Para analizar la robustez de los estimadores obtenidos, consideremos el caso de la media muestral en poblaciones normales: su varianza es σ^2/n, y puede demostrarse que este estimador es el más eficiente. Supongamos ahora una contaminación del tipo:

$$(1 - \alpha)N(\mu, \sigma) + \alpha N(\mu, k\sigma)$$

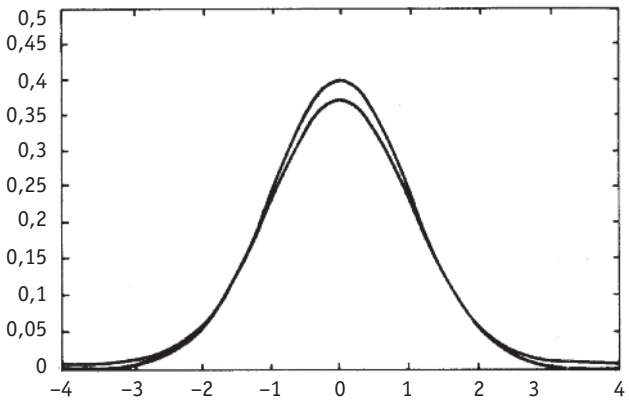
donde k es una constante arbitraria positiva (véase la figura 7.10); entonces, la media de la distribución mezclada sigue siendo μ pero su varianza resulta ser:

$$\sigma_c^2 = (1 - \alpha)\sigma^2 + ak^2\sigma^2 = \sigma^2(1 + \alpha[k^2 - 1])$$

Comparemos la precisión de $\bar{x}$ como estimador de μ en el modelo $N(\mu, \sigma)$ y en la normal contaminada. En ambos casos es centrado. Llamaremos $\bar{x}_1$ al estimador en el primer caso (es decir, a un estimador que sólo utiliza datos «buenos» generados por $N[\mu, \sigma]$) y $\bar{x}_2$ al estimador en el segundo caso. Entonces:

$$ER(\bar{x}_2/\bar{x}_1) = \frac{\text{Var}(\bar{x}_1)}{\text{Var}(\bar{x}_2)} = \frac{\sigma^2/n}{\sigma_c^2/n} = \frac{1}{1 + \alpha(k^2 - 1)}$$

Figura 7.10 Efecto de contaminar una $N(0, 1)$ con $\alpha = 0{,}1$, $k = 3$



Si, por ejemplo, $\alpha = 0{,}01$ y $k = 5$, $ER = 0{,}8$ y hay una pérdida de eficiencia del 20%, si $k = 7$ la pérdida pasa a ser del 35% y si aumentamos k manteniendo fijo α, la eficiencia tiende a cero.

La conclusión de este ejercicio es que una pequeña contaminación de la distribución, que suponga una baja probabilidad de generar datos muy heterogéneos, puede afectar drásticamente a la eficiencia del estimador.

Hay dos soluciones a este problema: la primera es utilizar estimadores robustos que, aunque no sean tan eficientes como los óptimos si el modelo es correcto, no cambien mucho sus propiedades ante contaminaciones como la estudiada. El segundo es utilizar los procedimientos clásicos y efectuar después un estudio de validación del modelo para identificar datos atípicos, como veremos en el capítulo 12.

Un compromiso razonable es calcular siempre junto al estimador clásico un estimador robusto: si ambos son análogos, tomar el clásico; si no lo son, someter a los datos al estudio exhaustivo de validación que se presenta en el capítulo 12.

7.5.6 Punto de ruptura de un estimador

Dada una muestra $\mathbf{X} = (x_1, ..., x_n)$, consideremos una nueva muestra ficticia $\mathbf{Y} = (y_1, \ldots y_n)$ construida sustituyendo m de los datos de $\mathbf{X}$ por valores arbitrarios. Por tanto $\mathbf{X}$ e $\mathbf{Y}$ tienen $n - m$ valores idénticos y m distintos. Dado un estimador $\hat{\vartheta}(\mathbf{X})$, llamaremos *alteración máxima* del estimador con contaminación m al valor máximo de la diferencia $|\hat{\vartheta}(\mathbf{X}) - \hat{\vartheta}(\mathbf{Y})|$ y escribiremos:

$$A(\mathbf{X}, m) = \max |\hat{\vartheta}(\mathbf{X}) - \hat{\vartheta}(\mathbf{Y})|$$

por ejemplo, para la media con $m = 1$ (modificando una única observación), $\hat{\vartheta}(\mathbf{Y})$ puede crecer sin límite y la alteración máxima es infinita. Definiremos *punto de ruptura* por:

$$\text{Punto de ruptura} = \frac{\{\text{máximo } m | \text{alteración máxima limitada}\}}{n}$$

El punto de ruptura de un estimador es pues la máxima fracción de la muestra que podemos cambiar sin causar un cambio arbitrario en el valor del estimador. Por ejemplo, cambiando un único dato podemos alterar a voluntad la media muestral, ya que, con n datos:

$$\bar{x} = \frac{(n-1)\bar{x}_{(n)} + x_n}{n}$$

donde hemos llamado $\bar{x}_{(n)}$ a la media de las $n - 1$ observaciones distintas de x_n. Esta expresión muestra que si fijamos n y $\bar{x}_{(n)}$ podemos alterar arbitrariamente $\bar{x}$ modificando x_n. El punto de ruptura de $\bar{x}$ es cero.

La mediana muestral es muy robusta: si con cinco datos hacemos arbitrariamente grandes o pequeños dos de ellos, la mediana tendrá un cambio controlado, ya que seguirá siendo uno de los tres datos muestrales no modificados. Su punto de ruptura es pues 2/5. En general, si n es impar podemos alterar arbitrariamente $(n - 1)/2$ datos sin llevarla fuera de los valores muestrales, por lo que el punto de ruptura es $1/2 - 1/2n$. Cuando n es par, el punto de ruptura es $1/2 - 1/n$. Por tanto, para n grande el punto de ruptura de la mediana es próximo a 0,5. Los estimadores robustos se construyen de manera que: (1) tengan punto de ruptura alto; (2) tengan una eficiencia razonable cuando los datos han sido generados por la distribución supuesta.

Medias recortadas

La media recortada a nivel α se calcula eliminando en la muestra el α% de las observaciones de cada extremo.

Por ejemplo, en una muestra de tamaño 10, la media recortada a nivel 0,2 (20%) es la media aritmética de las seis observaciones resultantes al eliminar las dos mayores y las dos menores. En general, si llamamos $x_{(i)}$ a los datos ordenados de manera que:

$$x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$$

llamando $m = n\alpha$ al número de observaciones eliminadas en cada extremo, que supondremos entero, la media recortada a nivel α, $T(\alpha)$, se calcula:

$$T(\alpha) = \frac{1}{n - 2m} \sum_{m+1}^{n-m} x_{(i)}$$

Los estudios realizados muestran que el grado óptimo de recorte es entre el 10 y el 25%. La pérdida de eficiencia con estos recortes es moderada. Por ejemplo, se demuestra que con $\alpha = 10\%$ la media recortada es un estimador centrado, con varianza

$$\text{Var}(T[10\%]) = 1{,}06\,\frac{\sigma^2}{n}$$

lo que supone sólo un 6% de pérdida de eficiencia con relación a la media muestral. El punto de ruptura de una media recortada es α.

Meda y variabilidad

La desviación típica tiene punto de ruptura cero y no es robusta. Una alternativa robusta es tomar la Meda muestral:

$$\text{Meda} = \text{mediana}\ \{|x_i - \ \text{Med}|\}$$

Esta estimación suele estandarizarse para que en poblaciones normales conduzca a un estimador consistente de σ. Para muestras grandes las desviaciones absolutas respecto a la mediana seguirán una distribución análoga a las desviaciones absolutas respecto a la media. De las tablas de la normal deducimos que el valor k que verifica

$$P(|x - \ \mu| \leq k\sigma) = 0{,}5$$

es $k = 0{,}675$. Como la Meda estima $k\sigma$, obtenemos un estimador consistente y robusto de σ en poblaciones normales con:

$$\hat{\sigma} = \frac{\text{Meda}}{0{,}675}$$

7.5.7 Propiedades de los estimadores por momentos

Los estimadores obtenidos por el método de los momentos son consistentes, pero no son, en general, ni centrados, ni con varianza mínima ni robustos. La ventaja de estos estimadores es su simplicidad; su inconveniente es que al no tener en cuenta la distribución de la población que genera los datos no utilizan toda la información de la muestra. El ejemplo 7.2 ilustraba esta situación. En la sección siguiente estudiaremos un procedimiento que proporciona estimadores con buenas propiedades, especialmente en muestras grandes: el método de máxima verosimilitud.

Ejercicios 7.3

7.3.1. Demostrar que cualquier combinación lineal $\sum\lambda_i\ \hat{\vartheta}_i$ de estimadores centrados para un parámetro es también centrada, si $\sum\lambda_i = 1$.

7.3.2. Para estimar la media de una población se considera el estimador $a \cdot \bar{x}$. Encontrar el valor de a que minimiza el error cuadrático medio de estimación.

7.3.3. Demostrar que la media muestral es un estimador consistente de la media de la población.

7.3.4. Obtener un estimador centrado para p en una distribución binominal y calcular su error cuadrático medio. ¿Es consistente?

7.3.5. Los defectos en una placa fotográfica siguen una distribución de Poisson. Se estudian siete placas encontrando 3, 5, 2, 1, 2, 3, 4 defectos. Encontrar un estimador centrado para λ, indicando la varianza del estimador.

7.3.6. Obtenga muestras, utilizando el método de Montecarlo, de una población normal (0, 1) y estudie la eficiencia relativa de la media y la mediana muestrales como estimadores de la esperanza de la distribución.

7.3.7. Demostrar que la media de dos observaciones cualesquiera en una muestra de tamaño n, ($n > 2$), es un estimador centrado para la media poblacional, pero no es consistente.

7.6 Estimadores de máxima verosimilitud

7.6.1 Introducción

El concepto de función de verosimilitud, debido a Fisher, es uno de los más importantes de la inferencia. Esta función se define partiendo de la distribución conjunta de la muestra, que se presenta a continuación.

7.6.2 La distribución conjunta de la muestra

Supongamos una variable discreta, x, con distribución $P(x, \vartheta)$ conocida. Al tomar muestras de tamaño n de esta población, cada muestra puede representarse por un vector $\mathbf{X}$, cuyos componentes son los valores observados. La distribución de este vector $\mathbf{X}$ cuando tomamos distintas muestras se denomina *distribución conjunta de la muestra*. Si la muestra es aleatoria simple, como:

$$P(\mathbf{X} = \mathbf{X}_0) = P(x_1 = x_{10}, x_2 = x_{20}, \ldots, x_n = x_{n0}) = P(x_{10}) \ldots P(x_{n0})$$

la probabilidad conjunta de la muestra es el producto de las probabilidades individuales. Por tanto, conociendo $P(x, \vartheta)$, podemos obtener fácilmente la probabilidad de cualquier muestra.

Cuando la variable sea continua, con función de densidad $f(x; \vartheta)$, la probabilidad del intervalo $x_1 - 1/2$, $x_1 + 1/2$, se aproxima por el rectángulo de altura $f(x_i)$ y base unidad:

$$P(x_i) = f(x_i) \cdot 1$$

Entonces, la probabilidad de la muestra será:

$$P(x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i)$$

Por tanto, la función de densidad conjunta de la muestra $f(x_1, ..., x_n)$ puede interpretarse, aproximadamente, como la probabilidad de obtener los valores muestrales $x_1 \pm 0{,}5, ..., x_n \pm 0{,}5$.

Ejemplo 7.6

Sea x una variable de Poisson con $\lambda = 2$. Calcular la probabilidad de obtener la muestra de tamaño cinco (3, 1, 0, 2, 0).

$$P(x_1 = 3, x_2 = 1, x_3 = 0, x_4 = 2, x_5 = 0) = P(3)P(1)P(0)P(2)P(0)$$

Como:

$$P(x) = \frac{e^{-2}2^x}{x!}$$

llamando:

$$\mathbf{X}_0' = (3 \;\; 1 \;\; 0 \;\; 2 \;\; 0)$$

$$P(\mathbf{X}_0) = \frac{e^{-2}2^3}{3!} \cdot \frac{e^{-2}2^1}{1!} \cdot \frac{e^{-2}2^0}{0!} \cdot \frac{e^{-2}2^2}{2!} \cdot \frac{e^{-2}2^0}{0!} =$$

$$= e^{-10} \cdot 2^6 \frac{1}{3!} \frac{1}{1!} \frac{1}{0!} \frac{1}{2!} \frac{1}{0!}$$

en general, llamando $x_1, ..., x_n$ a los valores muestrales, se obtiene:

$$P(\mathbf{X}) = e^{-n\lambda} \cdot \lambda^{\Sigma x_i} \frac{1}{\Pi x_i!}$$

que será la función de probabilidades conjunta. Nótese que todas las muestras que tengan iguales $\sum x_i$ y $1/x_i!$ tienen la misma probabilidad de ocurrir.

Ejemplo 7.7

Sea x binomial con $p = 0{,}2$ y $n = 10$. Calcular la probabilidad de obtener la muestra (1, 2, 1).

$$P(x_1=1,\, x_2=2,\, x_3=1)=\binom{10}{1}0{,}2^1\cdot 0{,}8^9\binom{10}{2}0{,}2^2\cdot 0{,}8^8\binom{10}{1}0{,}2^1\cdot 0{,}8^9=$$

$$=0{,}2^4\cdot\ 0{,}8^{26}\binom{10}{1}\binom{10}{2}\binom{10}{1}$$

En general, para una muestra $x_1 \ldots, x_k$

$$P(\mathbf{X}) = p^{\sum x_i}q^{nk-\sum x_i}\binom{n}{x_1}\ldots\binom{n}{x_k}$$

y todas las muestras que tengan el mismo $\sum x_i$, y los mismos valores de $\binom{n}{x_i}$ tendrán la misma probabilidad.

Ejemplo 7.8

Sea x exponencial de parámetro λ. Escribir la función de densidad conjunta de una muestra de tamaño n.

$$f(x) = \lambda e^{-\lambda x}$$

$$f(x_1, \ldots, x_n) = \Pi\lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda\sum x_i}$$

Nótese que aunque la función de densidad conjunta es en teoría n-dimensional, en este caso depende únicamente de $\sum x_i$ y, por tanto, todas las muestras que conduzcan al mismo valor $\sum x_i$ serán equiprobables.

7.6.3 La función de verosimilitud

Concepto

Supongamos una variable aleatoria continua x con función de densidad que representaremos por $f(x|\boldsymbol{\vartheta})$ para indicar que depende de un vector de parámetros $\boldsymbol{\vartheta}$, y una muestra aleatoria simple $\mathbf{X} = (x_1, ..., x_n)$. La función de densidad conjunta de la muestra es:

$$f(\mathbf{X} \mid \boldsymbol{\vartheta}) = \Pi f(x_i \mid \boldsymbol{\vartheta})$$

Cuando $\boldsymbol{\vartheta}$ es conocido, esta función determina la probabilidad de aparición de cada muestra.

En un problema de estimación se conoce un valor particular de $\mathbf{X}$, la muestra, pero $\boldsymbol{\vartheta}$ es desconocido. Sin embargo, la función anterior sigue siendo útil, ya que si sustituimos $\mathbf{X}$ por el valor observado, $\mathbf{X}_0 = (x_{10}, ..., x_{n0})$, la función $f(\mathbf{X}_0 \mid \boldsymbol{\vartheta})$ proporciona, para cada valor de $\boldsymbol{\vartheta}$, la probabilidad de obtener el valor muestral $\mathbf{X}_0$ para ese $\boldsymbol{\vartheta}$. Cuando variamos $\boldsymbol{\vartheta}$, manteniendo $\mathbf{X}_0$ fijo, se obtiene una función que llamaremos *función de verosimilitud*, $\ell(\boldsymbol{\vartheta}|\mathbf{X})$, o $\ell(\boldsymbol{\vartheta})$:

$$\ell(\boldsymbol{\vartheta}|\,\mathbf{X}) = \ell(\boldsymbol{\vartheta}) = f(\mathbf{X}_0 \mid \boldsymbol{\vartheta}) \qquad \mathbf{X}_0 \text{ fijo; } \boldsymbol{\vartheta} \text{ variable}$$

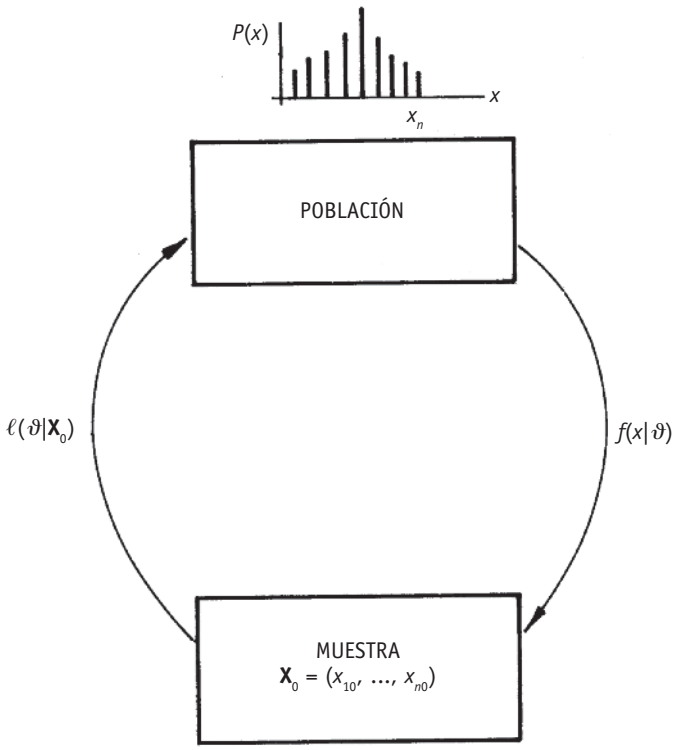
La figura 7.11 resume el concepto de verosimilitud: esta función aparece al invertir el papel de la función de densidad (o de probabilidades si la variable es discreta), consecuencia del cambio de óptica que tomamos en inferencia: en lugar de suponer que conocemos $\boldsymbol{\vartheta}$ y queremos calcular las probabilidades de distintas $\mathbf{X}$ posibles, suponemos que hemos observado una muestra $\mathbf{X}_0$ concreta —que se convierte por tanto en fija— y evaluamos la verosimilitud de los posibles valores de $\boldsymbol{\vartheta}$. Este cambio de perspectiva puede modificar la forma de la función completamente: si x es Poisson, y suponemos muestras de tamaño uno, la función de probabilidad de la muestra es:

$$P(x \mid \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$

y x toma únicamente valores discretos 0, 1, ... Al observar $x = 5$, la función de verosimilitud de esta muestra de tamaño uno será:

$$\ell(\lambda) = e^{-\lambda}\lambda^5$$

Figura 7.11 La función de verosimilitud



prescindiendo de la constante $(5!)^{-1}$; esta función es continua en λ y proporcional a la probabilidad de observar $x = 5$ para cada valor posible de λ.

Carácter no único de la verosimilitud

La función de verosimilitud se utilizará para comparar distintos valores del parámetro $\boldsymbol{\vartheta}$ dada la muestra. Si:

$$\ell(\boldsymbol{\vartheta}_1) = f(\mathbf{X}_0 \mid \boldsymbol{\vartheta}_1) > f(\mathbf{X}_0 \mid \boldsymbol{\vartheta}_2) = \ell(\boldsymbol{\vartheta}_2)$$

diremos que, a la vista de los datos muestrales, el valor $\boldsymbol{\vartheta}_1$ es más probable que el $\boldsymbol{\vartheta}_2$, ya que la probabilidad de obtener la muestra observada $\mathbf{X}_0$ es mayor con $\boldsymbol{\vartheta}_1$ que con $\boldsymbol{\vartheta}_2$.

Como la función de verosimilitud tiene unidades —las de medida de la variable x—, las diferencias entre verosimilitudes no tienen sentido, ya que pueden alterarse arbitrariamente al cambiar la escala de medida. Por ejemplo, supongamos que x está medida en metros, con función de densidad $f_x(x \mid \boldsymbol{\vartheta})$ e $y = 100x$ representa esta misma medida en cm. La función de densidad de y será:

$$f_y(\mathrm{y}) = \frac{1}{100} f_x\left(\frac{y}{100}\right)$$

La función de verosimilitud para $\boldsymbol{\vartheta}$ a partir de los datos en metros es:

$$\ell(\boldsymbol{\vartheta} \mid \mathbf{X}) = \Pi f_x(x_i \mid \boldsymbol{\vartheta})$$

Mientras que, con los datos en cm:

$$\ell(\boldsymbol{\vartheta} \mid \mathbf{Y}) = \Pi f_y(y_i \mid \boldsymbol{\vartheta}) = \left(\frac{1}{100}\right)^n \Pi f_x\left(\frac{y_i}{100} \mid \boldsymbol{\vartheta}\right) = \left(\frac{1}{100}\right)^n \ell(\boldsymbol{\vartheta} \mid \mathbf{X})$$

por tanto:

$$\ell(\boldsymbol{\vartheta}_1|\mathbf{Y}) - \ell(\boldsymbol{\vartheta}_2|\mathbf{Y}) = \left(\frac{1}{100}\right)^n [\ell(\boldsymbol{\vartheta}_1|\mathbf{X}) - \ell(\boldsymbol{\vartheta}_2|\mathbf{X})]$$

que muestra cómo las diferencias en verosimilitud se alteran arbitrariamente con la escala de medida. Por el contrario, los cocientes:

$$\frac{\ell(\boldsymbol{\vartheta}_1|\mathbf{Y})}{\ell(\boldsymbol{\vartheta}_2|\mathbf{Y})} = \frac{\ell(\boldsymbol{\vartheta}_1|\mathbf{X})}{\ell(\boldsymbol{\vartheta}_2|\mathbf{X})}$$

son invariantes.

En consecuencia, el valor absoluto de la verosimilitud es irrelevante, y sólo interesan las diferencias relativas. Por esta razón, si descomponemos la función de verosimilitud dada la muestra $\mathbf{X}_0$ en:

$$\ell(\boldsymbol{\vartheta}|\, \mathbf{X}_0) = g(\mathbf{X}_0) f(\mathbf{X}_0 \mid \boldsymbol{\vartheta})$$

donde $g(\mathbf{X}_0)$ es una función que depende sólo de los datos muestrales, es indiferente incluir o no esta función en la verosimilitud, ya que, dada la muestra, $g(\mathbf{X}_0)$ es constante y desaparecerá al comparar los cocientes de las verosimilitudes de dos valores posibles del parámetro.

La función soporte

En lugar del cociente, $\ell(\vartheta_2)|\ell(\vartheta_1)$, podemos usar la diferencia en logaritmos, $\ln\ell(\vartheta_2) - \ln\ell(\vartheta_1)$, para comparar los valores de la función de verosimilitud en distintos puntos. Al logaritmo de esta función:

$$\boxed{L(\vartheta) = \ln\ell(\vartheta)}$$

le llamaremos *función soporte y* no depende de constantes arbitrarias. Llamaremos *discriminación* contenida en la muestra **X** entre ϑ_2 y ϑ_1 a la diferencia de soporte de ambos valores. Si ϑ es un parámetro cuyos valores posibles pertenencen a un intervalo, llamaremos *discriminación relativa* entre ϑ_2 y ϑ_1 a:

$$\frac{L(\vartheta_2) - L(\vartheta_1)}{\vartheta_2 - \vartheta_1} = \frac{\ln\ell(\vartheta_2) - \ln\ell(\vartheta_1)}{\vartheta_2 - \vartheta_1}$$

y, en el límite, cuando ϑ_2 tiende a ϑ_1, obtendremos la *tasa de discriminación* de la muestra **X** respecto al parámetro ϑ en el punto ϑ_1:

$$d(\vartheta_1) = \lim_{\vartheta_2 \to \vartheta_1} \frac{L(\vartheta_2) - L(\vartheta_1)}{\vartheta_2 - \vartheta_1} = \frac{dL(\vartheta)}{d\vartheta}\bigg|_{\vartheta=\vartheta_1}$$

La tasa de discriminación, $d(\vartheta)$, introducida por Fisher, que la denominó «Score», juega un papel central en los procedimientos de inferencia. Intuitivamente vemos que si $d(\vartheta_1) > 0$, la verosimilitud aumenta para valores superiores a ϑ_1, es decir, la muestra tiene mayor probabilidad de ocurrir con valores mayores que ϑ_1, mientras que si $d(\vartheta_1) < 0$ el razonamiento se invierte.

Resumen

En resumen, la función de verosimilitud es la herramienta básica para juzgar la compatibilidad entre los valores muestrales observados y los posibles valores del parámetro. Para comparar dos posibles valores del parámetro debe utilizarse el cociente de sus verosimilitudes, y no su diferencia, que depende de la escala de medida de las variables.

Ejemplo 7.9

Para estimar el parámetro de la distribución de Poisson, la función de verosimilitud, dada la muestra $x_1, \ldots, x_n$, será:

$$\ell(\lambda) = \Pi P(x_i \mid \lambda) = e^{-n\lambda}\lambda^{\Sigma x_i}\frac{1}{\Pi x_i!}$$

Como el término $(\Pi x_i!)^{-1}$ no depende de λ, puede eliminarse si se desea y escribir la función como:

$$\ell(\lambda) = e^{-n\lambda}\lambda^{n\bar{x}}$$

que se representa en la figura 7.12(a). La función soporte será:

$$L(\lambda) = -n\lambda + n\bar{x} \ln \lambda$$

y tiene una estructura más simple.

Ejemplo 7.10

Para estimar la media y la varianza de una población normal

$$\ell(\mu, \sigma^2) = \Pi f(x_i \mid \mu, \sigma^2) = \frac{1}{\sigma^n} \frac{1}{(\sqrt{2\pi})^n} \exp\left[-\frac{1}{2\sigma^2} \Sigma(x_i - \mu)^2 \right]$$

y el soporte será:

$$L(\mu, \sigma^2) = -\frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \Sigma(x_i - \mu)^2$$

Utilizando la descomposición (7.6), podemos escribir:

$$\Sigma(x_i - \mu)^2 = \Sigma(x_i - \bar{x})^2 + n(\bar{x} - \mu)^2$$

con lo que la función soporte se convierte en:

$$L(\mu, \sigma^2) = -\frac{n}{2} \ln \sigma^2 - \frac{1}{2} \frac{ns^2}{\sigma^2} - \frac{n}{2\sigma^2} (\bar{x} - \mu)^2$$

Si σ^2 fuese conocido, la función se reduce a:

$$L(\mu) = k - \frac{n}{2\sigma^2} (\bar{x} - \mu)^2$$

y es una parábola con centro en $\bar{x}$ y curvatura en dicho punto $n/2\sigma^2$.

Las funciones de verosimilitud serán, la conjunta:

$$\ell(\mu, \sigma^2) = \sigma^{-n} e^{-ns^2/2\sigma^2} e^{-n(\bar{x}-\mu)^2/2\sigma^2}$$

Esta función depende de dos variables y no puede dibujarse fácilmente. Si suponemos σ^2 conocido, de manera que la función sólo dependa de μ, la función se simplifica a:

$$\ell(\mu) = k e^{-n(\bar{x}-\mu)^2/2\sigma^2}$$

donde k engloba todos los términos que dependen de σ^2. Esta función se representa en 7.8(b). Si suponemos μ conocido, el término $n(\bar{x} - \mu)^2/2$ se convierte en constante y, por lo tanto, el numerador del exponente de e será:

$$ns^2 + n(\bar{x} - \mu)^2 = \sum(x_i - \mu)^2$$

y la verosimilitud es:

$$\ell(\sigma^2) = (\sigma^2)^{-n/2} e^{-[\sum(x_i-\mu)^2]/2\sigma^2}$$

(véase la figura 7.12[c]).

Ejemplo 7.11

Supongamos que realizamos experimentos binomiales con tamaños de muestra $n_1, \ldots, n_k$ y contamos en cada uno de ellos el número de éxitos, $x_1, x_2, \ldots, x_k$. Si en todos ellos el parámetro p es constante, la función de verosimilitud será:

$$\ell(p) = \Pi p(x_i \mid n_i, p) = \Pi \binom{n_i}{x_i} \quad p^{x_i}(1-p)^{n_i - x_i} = p^{\sum x_i}(1-p)^{\sum n_i - \sum x_i} \Pi \binom{n_i}{x_i}$$

y, prescindiendo de constantes:

$$\ell(p) = p^{\sum x_i}(1-p)^{\sum n_i - \sum x_i}$$

representada en la figura 7.12(d).

Ejemplo 7.12

Dada la muestra $(x_1, \ldots, x_n)$ de una población uniforme $(0, b)$, la función de verosimilitud es:

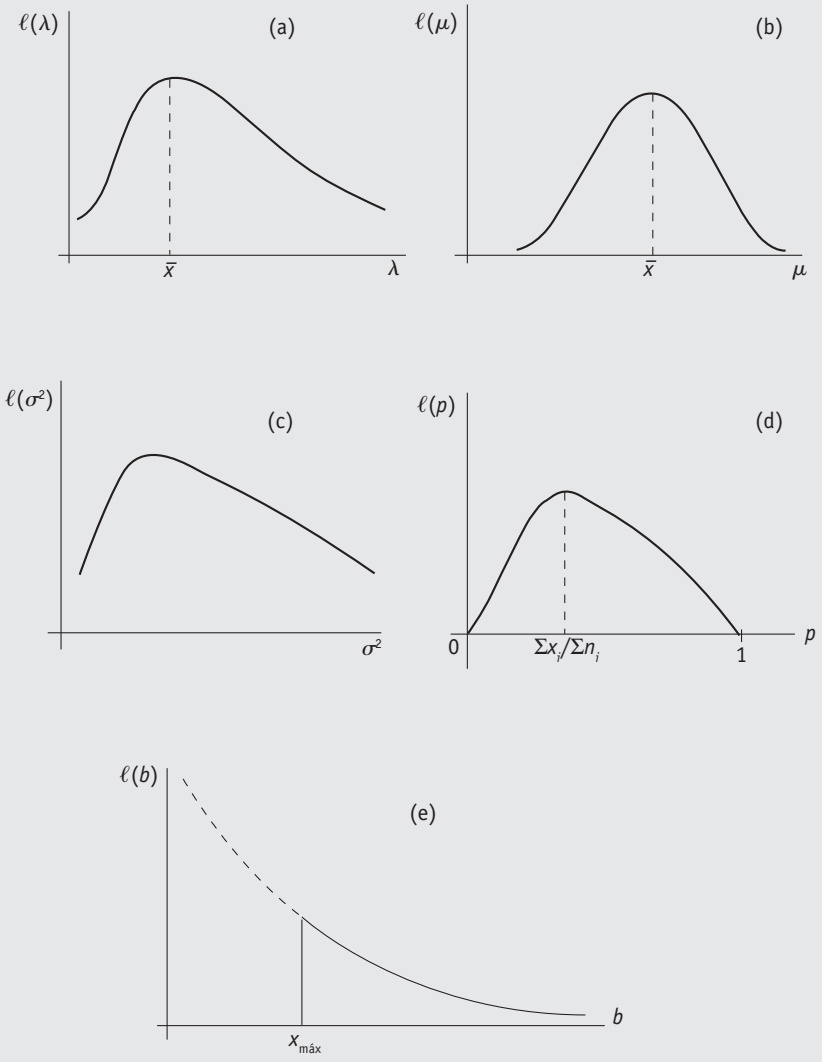
$$\begin{aligned} \ell(b) = \Pi f(x_i \mid b) &= 1/b^n \qquad \text{para } (x_1, \ldots, x_n \leq b) \\ &= 0 \qquad \text{en otro caso} \end{aligned}$$

Observemos que el valor del parámetro aparece también en el rango de definición de la función. La función de verosimilitud sólo es $1/b^n$ para b mayor que cualquiera de los valores observados. Es decir, considerando los valores muestrales $x_1, \ldots, x_n$ como números fijos y llamando $x_{\text{máx}}$ al mayor de todos ellos:

$$\ell(b) = 1/b^n \qquad b \geq x_{\text{máx}}$$

que se presenta en la figura 7.12(e).

Figura 7.12 **Funciones de verosimilitud para los parámetros siguientes: (a) λ en la distribución de Poisson; (b) μ en la normal con σ^2 conocido; (c) σ^2 en la normal; (d) p en la binomial; (e) b en la uniforme**



7.6.4 Estadísticos suficientes

Concepto

En los ejemplos anteriores la función de verosimilitud dependía de la muestra $\mathbf{X}$ únicamente a través de ciertas funciones $t(\mathbf{X})$. Por ejemplo, para

la distribución de Poisson la función de verosimilitud depende de la muestra solamente a través de $\sum x_i$, para la normal, a través de $\bar{x}$ y s^2, para la binomial a través de $\sum x_i$, y para la uniforme a través de $x_{\text{máx}}$. El conocimiento de estas funciones es pues suficiente para escribir la función de verosimilitud y, en este sentido, contienen toda la información de la muestra para estimar los parámetros correspondientes. Llamaremos estadísticos *suficientes* a estas funciones de los datos muestrales. En general, si la función de verosimilitud $\ell(\boldsymbol{\vartheta})$ para un vector de parámetros $\boldsymbol{\vartheta}$ de dimensión p depende de la muestra a través de ciertas funciones $t_1(\mathbf{X})$, ..., $t_h(\mathbf{X})$, con $h < n$, de los valores muestrales, diremos que estos estadísticos son suficientes para $\boldsymbol{\vartheta}$. En ciertos casos, como ocurre en los ejemplos anteriores, $h = p$, pero puede no ser así.

Los estadísticos suficientes no son únicos, ya que si $\sum x_i$ es suficiente para λ en un modelo de Poisson, también lo será $\sum x_i/n = \bar{x}$ o $\sum x_i/8$. Algunas de estas funciones tendrán buenas propiedades como estimadores del parámetro y entonces las llamaremos estimadores suficientes.

No conviene olvidar que los estadísticos suficientes sólo contienen toda la información respecto al parámetro cuando el modelo supuesto es cierto. Sin embargo, para contrastar esta hipótesis serán necesarios todos los valores muestrales. Por ejemplo, en una población normal, con σ conocida, $\bar{x}$ es suficiente para estimar μ, es decir, para estimar la media de la población sólo necesitamos el valor de la media muestral y los valores particulares de la muestra son irrelevantes. Sin embargo, para contrastar la normalidad de los datos son necesarios todos los valores muestrales observados, como veremos en el capítulo 12.

Es intuitivo que los estimadores basados en estadísticos suficientes serán, en algún sentido, «óptimos» al utilizar toda la información, y que su existencia dependerá del modelo supuesto. Este último aspecto se precisa en la condición siguiente.

Criterio de Fisher-Neyman

Para saber si existen estadísticos suficientes para un parámetro $\boldsymbol{\vartheta}$ dentro del modelo f estudiaremos si la función de densidad conjunta puede descomponerse en:

$$f(\mathbf{X}|\boldsymbol{\vartheta}) = g(\mathbf{t}[\mathbf{X}]|\boldsymbol{\vartheta})h(\mathbf{X})$$

donde $\mathbf{t}(\mathbf{X})$ es un vector de funciones de los valores muestrales. Sus componentes serán estadísticos suficientes, ya que, al conocerlos, podemos escribir la función de verosimilitud como:

$$\ell(\boldsymbol{\vartheta}) = g(\mathbf{t}[\mathbf{X}]|\boldsymbol{\vartheta})$$

Este criterio es debido a Fisher y Neyman.

7.6.5 El método de máxima verosimilitud

Supongamos construida la función de verosimilitud para $\boldsymbol{\vartheta}$, $\ell(\boldsymbol{\vartheta})$. Un procedimiento intuitivo de estimación es escoger aquel valor que haga máxima la probabilidad de aparición de los valores muestrales efectivamente observados; en otros términos, seleccionar como estimador del parámetro el valor que maximice la probabilidad de lo efectivamente ocurrido. Esto conduce a obtener el valor máximo de la función $\ell(\boldsymbol{\vartheta})$. Suponiendo que esta función es diferenciable y que su máximo no ocurre en un extremo de su campo de definición, el máximo se obtendrá resolviendo el sistema de ecuaciones:

$$\frac{\partial \ell(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}_1} = 0$$
$$\cdots\cdots\cdots\cdots$$
$$\frac{\partial \ell(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}_p} = 0$$

El valor resultante así obtenido, $\hat{\boldsymbol{\vartheta}}$, corresponderá a un máximo si la matriz hessiana de segundas derivadas H, evaluada en dicho punto $\hat{\boldsymbol{\vartheta}}$, es definida negativa:

$$H(\hat{\boldsymbol{\vartheta}}) = \left(\frac{\partial^2 \ell[\boldsymbol{\vartheta}]}{\partial \boldsymbol{\vartheta}_i \partial \boldsymbol{\vartheta}_j} \right)_{\vartheta = \hat{\vartheta}} \quad \text{definida negativa.}$$

En la práctica, los estimadores máximo-verosímiles (*MV*) se obtienen derivando en el logaritmo de la función de verosimilitud o *función soporte:*

$$L(\boldsymbol{\vartheta}) = \ln \ell(\boldsymbol{\vartheta})$$

Como el logaritmo es una transformación monótona, las funciones soporte, $L(\boldsymbol{\vartheta})$, y verosimilitud, $\ell(\boldsymbol{\vartheta})$, tendrán el mismo máximo. El soporte tiene la ventaja de que al tomar logaritmos las constantes multiplicativas se hacen aditivas y desaparecen al derivar, con lo que la derivada del soporte tiene siempre la misma expresión y no depende de constantes arbitrarias, como ocurre con la derivada de la verosimilitud.

La derivada de la función soporte es la tasa de discriminación, y podemos definir el estimador máximo-verosímil como aquel valor del parámetro para el que se anula la tasa de discriminación de la muestra.

Ejemplo 7.13

Obtener en una población normal los estimadores *MV* de μ y σ^2.

El soporte de la muestra es (ejemplo 7.10):

$$L(\mu, \sigma^2) = -\frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \Sigma(x_i - \mu)^2$$

Para obtener los estimadores de máxima verosimilitud derivaremos e igualaremos a cero; entonces:

$$\frac{\partial L}{\partial \mu} = 0 = \frac{1}{\sigma^2} \Sigma(x_i - \mu) = \frac{n\bar{x} - n\mu}{\sigma^2}$$

$$\frac{\partial L}{\partial \sigma^2} = 0 = -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \Sigma(x_i - \mu)^2$$

La resolución de este sistema de ecuaciones conduce a:

$$\hat{\mu} = \bar{x}$$

$$\hat{\sigma^2} = \frac{\Sigma(x_i - \bar{x})^2}{n} = s^2$$

que son los estimadores de máxima verosimilitud. Comprobemos que la matriz hessiana es, para estos valores, definida negativa:

$$\frac{\partial^2 L}{\partial \mu^2} = -\frac{n}{\sigma^2}$$

$$\frac{\partial^2 L}{\partial \mu \partial \sigma^2} = \frac{-n(\bar{x} - \mu)}{\sigma^4}$$

$$\frac{\partial^2 L}{(\partial \sigma^2)^2} = +\frac{n}{2} \frac{1}{\sigma^4} - \frac{\Sigma(x_i - \mu)^2}{\sigma^6}$$

Particularizando estas derivadas en el máximo $(\bar{x}, s^2)$, la matriz hessiana será:

$$H(\bar{x}, s^2) = \begin{vmatrix} -\frac{n}{s^2} & 0 \\ 0 & \frac{-n}{2s^4} \end{vmatrix}$$

7.6.6 Propiedades de los estimadores máximo-verosímiles

Para distribuciones cuyo rango de valores posibles es conocido a priori y *no depende de ningún parámetro* (observemos que esto excluye la distribución uniforme [*0*, *b*]), puede demostrarse que, en condiciones muy generales respecto al modelo de distribución de probabilidad, el método de máxima verosimilitud proporciona estimadores que son:

1) Asintóticamente centrados.
2) Con distribución asintóticamente normal.
3) Asintóticamente de varianza mínima (eficientes).
4) Si existe un estadístico suficiente para el parámetro, el estimador máximo-verosímil es suficiente.
5) Invariantes en el sentido siguiente: si $\hat{\vartheta}_{MV}$ es el estimador máximo-verosímil de un parámetro ϑ, y g es una función cualquiera, $g(\hat{\vartheta}_{MV})$ es el estimador máximo-verosímil de $g(\vartheta)$.

Comentemos estas propiedades.

Asintóticamente centrados

Ésta es una propiedad muy general de estimadores razonables. Típicamente el sesgo de un estimador decrece con n (véase, por ejemplo, [7.7] o [7.17]), por lo que para muestras grandes el estimador es prácticamente centrado. Sin embargo, para pequeñas muestras los estimadores *MV* son frecuentemente sesgados.

Asintóticamente normales

Para tamaños muestrales grandes, si desarrollamos en serie la función soporte en un entorno del estimador *MV*, $\hat{\vartheta}_{MV}$, tendremos:

$$L(\vartheta) \simeq L(\hat{\vartheta}_{MV}) + \frac{1}{2}\left(\frac{d^2L[\hat{\vartheta}_{MV}]}{d\vartheta^2}\right)(\vartheta - \hat{\vartheta}_{MV})^2$$

que es una función cuadrática. Llamando:

$$\boxed{\hat{\sigma}^2_{MV} = \left[-\frac{d^2L(\hat{\vartheta}_{MV})}{d\vartheta^2}\right]^{-1}} \qquad (7.22)$$

la función de verosimilitud resultante puede escribirse:

$$\ell(\vartheta|\mathbf{X}) = k \exp\left\{-\frac{1}{2\hat{\sigma}^2_{MV}}(\vartheta - \hat{\vartheta}_{MV})^2\right\}$$

La relación entre la función de densidad conjunta y la verosimilitud es $f(\mathbf{X}|\vartheta) = h(\mathbf{X})\ell(\vartheta|\mathbf{X})$, donde $h(\mathbf{X})$ es una función sólo de los datos muestrales y podemos escribir

$$f(\mathbf{X}|\vartheta) = h(\mathbf{X}) \exp\left\{-\frac{1}{2\hat{\sigma}^2_{MV}}(\hat{\vartheta}_{MV} - \vartheta)^2\right\} \tag{7.23}$$

La función de densidad conjunta puede descomponerse de acuerdo con el criterio de Fisher-Neyman y por tanto el estimador $\hat{\vartheta}_{MV}$ es suficiente. Entonces la función de densidad conjunta de las observaciones puede escribirse como $f(\mathbf{X}|\vartheta) = f(\mathbf{X}|\hat{\vartheta}_{MV})f(\hat{\vartheta}_{MV}|\vartheta)$, donde el primer término representa la distribución de las observaciones dado el estadístico suficiente $\hat{\vartheta}_{MV}$ y no depende del parámetro y el segundo término representa la distribución del estadístico. Deducimos que en (7.23) el segundo término proporciona, salvo constantes, la densidad $f(\hat{\vartheta}_{MV}|\vartheta)$ y el estimador $\hat{\vartheta}_{MV}$ tendrá una distribución normal con media ϑ y varianza asintótica (7.22). Este resultado implica, además, que podemos calcular siempre la varianza asintótica del estimador *MV* mediante (7.22).

Para interpretar la expresión (7.22) observemos que la segunda derivada del soporte en el máximo es su curvatura. Si ésta es grande (figura 7.13 [b]) el máximo está bien definido y variará poco de muestra en muestra. Cuando la curvatura sea pequeña, la función es plana en el máximo (figura 7.13 [a]), y pequeñas variaciones muestrales modificarán mucho su posición. Es intuitivo que la varianza del estimador —que es la variabilidad del máximo de $L(\theta)$ en distintas muestras— es inversamente proporcional a la curvatura observada. Fisher denominó *información observada* a la segunda derivada del soporte en el máximo cambiada de signo, que es, según (7.22), la inversa de la varianza asintótica del estimador *MV* (en el apéndice 7B se detallan estas ideas).

Cuando $\boldsymbol{\vartheta}$ es vectorial, la matriz de varianza y covarianzas, $\mathbf{Var}(\hat{\boldsymbol{\vartheta}}_{MV})$, verifica:

$$\mathbf{Var}(\hat{\boldsymbol{\vartheta}}_{MV}) \rightarrow \left[-\frac{\partial^2 L(\hat{\boldsymbol{\vartheta}}_{MV})}{\partial\vartheta_i\partial\vartheta_j}\right]^{-1} = -\mathbf{H}(\hat{\boldsymbol{\vartheta}}_{MV})^{-1}$$

donde $\mathbf{H}$ es la matriz de segundas derivadas evaluada en el punto $\hat{\boldsymbol{\vartheta}}_{MV}$. Llamaremos matriz de información observada a:

$$\mathbf{IO}(\hat{\boldsymbol{\vartheta}}) = -\mathbf{H}(\hat{\boldsymbol{\vartheta}}_{MV})$$

y se verifica que, asintóticamente:

$$\mathbf{Var}(\hat{\boldsymbol{\vartheta}}_{MV})^{-1/2}(\hat{\boldsymbol{\vartheta}}_{MV} - \boldsymbol{\vartheta}) \rightarrow N(\mathbf{0}, \mathbf{I})$$

Asintóticamente eficientes

En el apéndice 7B se explica con detalle que, en condiciones muy generales, existe una cota mínima a la varianza de cualquier estimador centrado en poblaciones regulares. Esta cota es:

$$\mathrm{Var}(\hat{\vartheta}) \geq \frac{1}{E\left[-\dfrac{d^2L(\vartheta)}{d\vartheta^2}\right]_{\theta=\theta_V}} \tag{7.24}$$

y el denominador se llama información esperada, donde θ_V es el verdadero valor del parámetro. Para tamaños muestrales grandes, la información observada (7.22) converge hacia la información esperada y el estimador *MV* es óptimo.

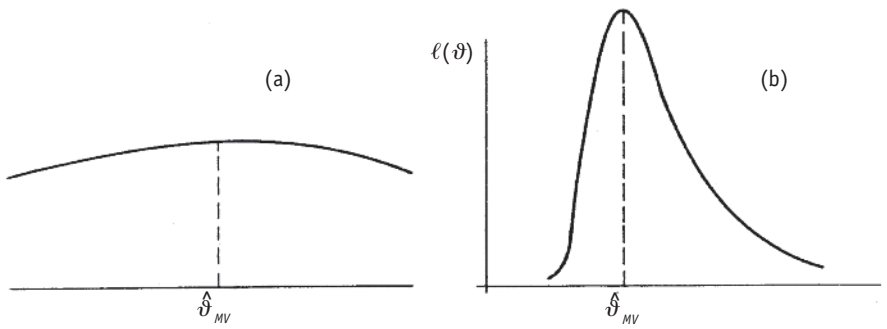
Este resultado se generaliza fácilmente al caso vectorial (apéndice 7B).

Suficiencia

Si existe un estimador suficiente, $t(\mathbf{X})$, la función de verosimilitud se escribe:

$$\ell(\vartheta) = g(\vartheta|t[\mathbf{X}]) \cdot h(\mathbf{X})$$

Figura 7.13 La función de verosimilitud



al derivar e igualar a cero, la solución de:

$$\frac{dg(\vartheta|t[\mathbf{X}])}{d\vartheta} = 0$$

será forzosamente función de $t(\mathbf{X})$ y, por tanto, suficiente.

Invarianza

Ésta es una propiedad muy útil porque permite obtener el estimador *MV* de cualquier función del parámetro. Por ejemplo si $\bar{x}$ es *MV* para μ, los estimadores *MV* de $\ln \mu$ o μ^2 son respectivamente $\ln \bar{x}$ o $\bar{x}^2$.

La demostración para el caso más simple en que la función tiene inversa única es la siguiente: sea ϑ_{MV} el valor que hace cero la derivada de la función soporte $L'(\hat{\vartheta}_{MV}) = 0$. Sea $\beta = g(\vartheta)$ una función del parámetro. Entonces la verosimilitud será:

$$L(g^{-1}[\beta])$$

Derivando respecto a β y aplicando la regla de la cadena:

$$\frac{\partial L(g^{-1}[\beta])}{\partial (g^{-1}[\beta])} \cdot \frac{\partial g^{-1}(\beta)}{\partial \beta}$$

El valor $\hat{\beta} = g(\hat{\vartheta}_{MV})$ hace cero el primer término y, por tanto, es un estimador *MV* para β.

Ejemplo 7.14

Obtener la varianza asintótica de los estimadores $\bar{x}$ y s^2 en una población normal y comparar con los resultados exactos.

Con los resultados del ejemplo 7.13 la matriz de varianzas asintóticas será:

$$\mathbf{Var}(\bar{x}, s^2) = \begin{vmatrix} s^2/n & 0 \\ 0 & \dfrac{2s^4}{n} \end{vmatrix}$$

Comparando con los resultados exactos, vemos que la varianza asintótica de $\bar{x}$ es s^2/n y la exacta σ^2/n. Respecto a s^2, la varianza asintótica es $2s^4/n$ y la exacta $2\sigma^4/(n-1)$. Como s^2 converge a σ^2, para muestras grandes los resultados serán análogos. Observemos que las covarianzas son nulas.

Robustez

Los estimadores *MV* no son en general robustos, y una sola observación atípica puede afectar mucho a sus propiedades. Conviene antes de calcularlos realizar el análisis exporatorio de datos estudiado en los capítulos 2 y 3 para asegurarse de que no existen valores atípicos extremos que pueden distorsionar el cálculo del estimador.

Ejercicios 7.4

7.4.1. Una máquina puede averiarse por dos razones, *A* y *B*. Se desea estimar la probabilidad de avería diaria de cada tipo sabiendo que:
   a) La probabilidad de avería de tipo *A* es doble que la de *B*.
   b) No existen otros tipos de averías posibles.
   c) Se han observado 30 días con el resultado siguiente: dos averías tipo *A*; tres tipo *B*; 25 días sin avería.

7.4.2. Dada la variable con densidad $f(x) = 2 \cdot \vartheta^{-2}(\vartheta - x)$, $0 < x \leq \vartheta$, calcular un estimador máximo-verosímil para ϑ.

7.4.3. Si $f(x) = \vartheta x^{\vartheta-1}$ $(0 < x < 1)$, encontrar un estadístico suficiente para ϑ y el estimador máximo-verosímil de ϑ, calculando su varianza asintótica.

7.4.4. Si $f(x) = 1/(\sqrt{2\pi}\,|\vartheta|)\exp(-[x-\vartheta]^2/2\vartheta^2)$, encontrar un estadístico suficiente para ϑ y el estimador máximo-verosímil.

7.4.5. Si x es log normal con densidad $f(x) = 1/(x\sigma\sqrt{2\pi})\exp -(\ln x - \mu)^2/2\sigma^2$ para $x > 0$, encontrar el estimador máximo-verosímil de μ y σ^2. Indicar si estos estimadores son o no funciones de los estadísticos suficientes.

7.4.6. Encontrar estimadores de máxima verosimilitud para los parámetros (a, b) en una distribución uniforme entre dichos valores.

7.4.7. Obtener el estimador por momentos y el de *MV* para el parámetro en la distribución de Pareto $f(x) = (\vartheta/x_0)(x_0/x)^{\vartheta+1}$ $(x \geq x_0 > 0;\ \vartheta > 0)$.

7.4.8. El control de recepción de una partida de rodillos se realiza clasificando las piezas en pequeñas, normales y grandes. Las proporciones teóricas se suponen $p_1 = 0{,}05$, $p_2 = 0{,}90$, $p_3 = 0{,}05$, pero se sospecha que ha aumentado la dispersión y, por tanto, las piezas anormales, según $p_1 = 0{,}05 + \vartheta$, $p_2 = 0{,}90 - 2\vartheta$, $p_3 = 0{,}05 + \vartheta$. Se analizan 5.000 piezas obteniendo $n_1 = 278$; $n_2 = 4.428$; $n_3 = 294$ de cada clase. Se pide la estimación *MV* de ϑ.

7.4.9. Estimar por el método *MV* el parámetro p en la ley geométrica $p_k = pq^{k-1}$. Calcular la varianza asintótica del estimador.

7.4.10. Calcular el estimador *MV* de ϑ en la función $f(x) = \vartheta(1-x)^{\vartheta-1}$ $(0 < x < 1)$.

7.4.11. Para determinar la vida media de unos componentes se selecciona una muestra de 10 unidades, con los resultados siguientes: 20, 50, 80, 40, 25, 85. El ensayo se detiene al cabo de 85 horas y, para entonces, cuatro unidades seguían en funcionamiento. Admitiendo que la distribución de vida es exponencial, estimar la vida media de estos componentes.

7.4.12. Los elementos de un proceso de fabricación pueden tener cualquier combinación de los defectos A, B, C y D, aunque cada defecto sólo puede darse una vez en cada elemento. Se tiene $p(A) = p(B) = p(C) = p_1$; $p(D) = p_2$. Se pide:

a) Calcular la distribución de probabilidad del número total de defectos en un elemento.
b) Calcular la media y varianza de la distribución.
c) Para estimar p_1 y p_2 se toma una muestra de 200 elementos, encontrando 12 con sólo el defecto A, 8 con sólo el B, 10 con el C y 18 con el D. Estimar p_1 y p_2 por el método de máxima verosimilitud.

7.4.13. Demostrar que la constante a en $a\sum(x_i - x)^2$ que minimiza el error cuadrático medio de estimación de σ^2 es $a = 1/(n+1)$ (suponer normalidad).

7.4.14. Demostrar que s^2 es consistente para σ^2 en poblaciones normales.

7.4.15. Obtener el estimador *MV* del cociente señal/ruido (μ/σ) en una población normal.

7.4.16. Un taller dispone de dos tipos de máquinas: el primero produce componentes con resistencia eléctrica media μ_1 y desviación σ_1; el segundo, con media μ_2 y desviación σ_2. Los elementos se mezclan aleatoriamente en la fabricación final, pero se conoce que el 40% proviene de la primera máquina y el 60% de la segunda. Calcular μ y σ para la distribución final de la fabricación, suponiendo normalidad.

7.4.17. Si k es un número real positivo, para una población normal:

a) Obtener el estimador *MV* de σ^{2k}.
b) Estudiar si este estimador es centrado, sabiendo que $E[\chi_n^{2k}] = h(n) \neq n^k$.
c) Si no lo es, obtener uno que lo sea.

7.7 Resumen del capítulo y consejos de cálculo

En este capítulo se ha presentado el concepto básico de distribuciones en el muestreo y el cuadro 7.2 resume estas distribuciones para algunos estadísticos importantes. La distribución en el muestreo de un estimador describe sus propiedades principales y sirve para comparar estimadores. Es deseable que el estimador esté centrado en el valor del parámetro y tenga mínima variabilidad. Podemos comparar estimadores con el criterio de error cuadrático medio, que es el cuadrado del error promedio cometido al estimar el parámetro mediante el estimador. Un método general para obtener estimadores es el método de máxima verosimilitud, que proporciona estimadores con buenas propiedades en muestras grandes. Los estimadores de máxima verosimilitud son muy sensibles a datos atípicos, por lo que conviene limpiar la muestra de estos datos antes de aplicar el procedimiento.

Podemos obtener la distribución en el muestreo de cualquier estadístico mediante el método de Montecarlo con cualquier programa estadístico, incluyendo Excel, como comentamos en el capítulo 5. La maximización de la verosimilitud en problemas más complejos que los aquí estudiados requiere algoritmos de optimización de funciones que se encuentran disponibles en los paquetes estadísticos habituales.

Cuadro 7.2 Resumen de distribuciones en el muestreo

Población	Estadístico	Distribución	Media	Desviación típica
Binomial (p, n)	$\hat{p}$	Aprox. normal (n grande)	p	$\sqrt{pq/n}$
Cualquiera (μ, σ)	$\bar{x}$	Aprox. normal (n grande)	μ	$\sigma/\sqrt{n}$
Normal (μ, σ)	$\bar{x}$	Normal	μ	$\sigma/\sqrt{n}$
Normal (μ, σ)	s^2	$\sigma^2\chi^2_{n-1}/n$	$(n-1)\sigma^2/n$	$\sigma^2\sqrt{2(n-1)/n}$
Normal (μ, σ)	$\hat{s}^2$	$\sigma^2\chi^2_{n-1}/(n-1)$	σ^2	$\sigma^2\sqrt{2(n-1)}$
Normal (μ, σ)	$\hat{s}$	$\sigma\chi_{n-1}/\sqrt{n-1}$	$\frac{(4n-5)}{4n-4}\sigma$	$\sigma^2\sqrt{2(n-1)}$
Cualquiera	$\hat{s}^2$	—	σ^2	$\sigma^2\sqrt{\frac{2}{n-1}+\frac{CA_p-3}{n}}$

7.8 Lecturas recomendadas

Todos los manuales de estadística básica que se listan en la bibliografía incluyen capítulos de estimación por punto y por intervalo. Lehmann y Casella (2003) es un tratamiento riguroso, Lindgren (1993) y Guttman *et al.* (1982) son especialmente claros, y Rohatgi (1976) incluye numerosos ejemplos. Silvey (1970) es una excelente aunque condensada presentación de estos conceptos. A nivel más simple, Freedman, Pisani y Purves (2007), Wonnacott y Wonnacott (2004) y Newbold *et al.* (2006). Larsen y Marx (2005) es un texto muy recomendable, con una excelente colección de datos reales y un nivel matemático algo superior al de este libro. Este capítulo se basa en el muestreo aleatorio simple. Para otros tipos de muestreo, véanse Azorín y Sánchez Crespo (1986), Mirás (1985) y Cochran (1980).

Apéndice 7A: Muestreo en poblaciones finitas

Cuando el tamaño de la población (N) es pequeño con relación a la fracción estudiada (n), la distinción entre muestreo con y sin reemplazamiento es importante. Suponiendo no reemplazamiento, la media muestral ($\bar{x}$) es todavía un estimador centrado de μ. Sin embargo, su varianza es ahora menor que σ^2/n. Para calcularla tenemos que tener en cuenta que ahora las x_i son dependientes (debido al no reemplazamiento) y:

$$\text{Var}(\bar{x}) = \text{Var}\left[\frac{1}{n}(x_1 + ... + x_n)\right] =$$

$$= \frac{1}{n^2}\sum \text{Var}(x_i) + \frac{2}{n^2}\binom{n}{2}\text{Cov}(x_i x_j) \qquad (7A.1)$$

Para calcular la covarianza entre dos observaciones cualesquiera aplicamos la definición:

$$\text{Cov}(x_i x_j) = \sum_{i=1}^{N}\sum_{\substack{j\neq i\\ j=1}}^{N}(x_i - \mu)(x_j - \mu)\frac{1}{N(N-1)} \qquad (7A.2)$$

donde $N(N-1)$ es el número de términos que sumamos. Como:

$$\sum_{\substack{j\neq i\\ j=1}}^{N}(x_j - \mu) = -(x_i - \mu) \qquad (7A.3)$$

sustituyendo en (7A.2):

$$\text{Cov}\,(x_i x_j) = -\sum_{i=1}^{N}(x_i - \mu)^2 \frac{1}{N(N-1)} = -\frac{\sigma^2}{N-1} \qquad (7A.4)$$

ya que, por definición, la varianza de la población es $\sum(x_i - \mu)^2/N$. Sustituyendo (7A.4) en (7A.1):

$$\text{Var}\,(\bar{x}) = \frac{n\sigma^2}{n^2} + \frac{2}{n^2}\frac{n(n-1)}{2}\left(-\frac{\sigma^2}{N-1}\right) = \frac{\sigma^2}{n}\left[\frac{N-n}{N-1}\right] \qquad (7A.5)$$

El término $(N-n)/(N-1)$ se denomina *factor de corrección en poblaciones finitas*, este factor se escribe también:

$$f = \frac{1-\frac{n}{N}}{1-\frac{1}{N}}$$

y cuando n/N es pequeño (menor de 0,1) y N mediano (mayor de 30) este término es prácticamente la unidad.

La conclusión fundamental de este resultado es que la precisión de $\bar{x}$ para estimar μ (y, como caso particular, de $\hat{p}$ para estimar p) depende sólo de n, y no del tamaño de la población siempre que n/N sea pequeño y N moderadamente grande.

Apéndice 7B: Estimadores eficientes, el concepto de información

Llamaremos estimador *eficiente* a aquel que es centrado y tiene varianza mínima. Estos estimadores se reconocen mediante la cota de Cramer-Rao, que establece una cota mínima para la varianza de cualquier estimador de un parámetro ϑ, en un modelo que verifique ciertas condiciones generales de regularidad.

La más importante de estas condiciones es que el rango de variación de la variable no dependa del parámetro a estimar. Por tanto, la distribución uniforme $(0, \vartheta)$ no es regular para estimar ϑ. Las distribuciones binomial, Poisson, normal y sus distribuciones asociadas son todas regulares.

La cota de Cramer-Rao establece que la varianza de cualquier estimador centrado de ϑ, $\hat{\vartheta}_c$ debe verificar:

$$\text{Var}\,(\hat{\vartheta}_c) \geq E\left[-\left(\frac{d^2L[\vartheta]}{d\vartheta^2}\right)_{\vartheta_v}\right]^{-1}$$

donde la segunda derivada de la función soporte está evaluada en ϑ_v, valor verdadero del parámetro.

Para interpretar este resultado, observemos que en la función soporte:

a) La derivada es cero para $\vartheta = \hat{\vartheta}_{MV}$, estimador máximo-verosímil. En el valor verdadero ϑ_v esta derivada podrá ser negativa o positiva, dependiendo de la muestra.
b) La segunda derivada es proporcional a la curvatura de la función.

Para cualquier función, $f(x)$, la curvatura en un punto (que es la inversa del radio del círculo que mejor aproxima la función en dicho punto) es:

$$C(x) = \text{curvatura}\,(x) = \frac{f''(x)}{(1 + f'[x]^2)^{3/2}}$$

como en el máximo de la función soporte, $\vartheta = \hat{\vartheta}_{MV}$, la primera derivada es nula, la segunda derivada:

$$\frac{d^2L}{d\vartheta^2}$$

representa la curvatura en ese punto. Por tanto, cuando la segunda derivada sea grande, la muestra apunta muy claramente hacia el valor del parámetro $\vartheta = \hat{\vartheta}_{MV}$, mientras que si la curvatura es débil, hay un conjunto grande de valores del parámetro que conducen casi al mismo valor de la función soporte y son estimaciones del parámetro dada la muestra casi igualmente razonables.

El cuadro 7.3 representa $L(\vartheta)$ y sus dos primeras derivadas para algunos casos simples uniparamétricos. Se observa que la segunda derivada aumenta con n; es decir, la curvatura de la función soporte aumenta con el tamaño muestral, como sería de esperar.

Establecidas estas propiedades, consideremos lo que ocurre cuando tomamos muchas muestras y analizamos las propiedades *promedio* de las dos primeras derivadas de la función soporte, medidas en el valor verdadero del parámetro, ϑ_v.

La tasa de discriminación, $dL(\vartheta)/d\vartheta$, para $\vartheta = \vartheta_v$ no será, en general, cero: el cuadro 7.3 muestra que, por ejemplo, al estimar la media de una población normal, la tasa de discriminación será negativa o positiva según que $\bar{x}$ sea menor o mayor que ϑ_v. Sin embargo, se observa en los cinco ejemplos que su valor promedio es cero. Este resultado puede demostrarse

en condiciones muy generales para aquellas distribuciones que tienen un rango de variación que no depende de ningún parámetro (como la normal, exponencial, Poisson, etc.; obsérvese que la uniforme [0, ϑ] no cumple esta condición).

$$E\left[\frac{dL(\vartheta)}{d\vartheta}\right]_{\vartheta=\vartheta_v} = 0$$

En términos poco precisos, esta expresión indica que una muestra aleatoria por término medio indica correctamente el valor del parámetro.

Consideremos qué ocurre ahora al valor esperado de la segunda derivada, proporcional a la curvatura: el cuadro 7.3 muestra que, en primer lugar, su valor aumenta con n, indicando mayor precisión al aumentar el tamaño muestral; en segundo lugar su valor puede ser constante y no depender del valor del parámetro, como en el caso 4, o ser función de éste (casos 1, 2, 3 y 5). Intuitivamente, cuanto mayor sea la curvatura promedio de $L(\vartheta)$, más precisa puede ser la estimación del parámetro.

La cota de Cramer-Rao establece que la varianza mínima de un estimador centrado de ϑ depende del radio de curvatura esperado.

Fisher denominó a la cantidad:

$$\boxed{IE(\vartheta) = E\left[-\frac{d^2L(\vartheta)}{d\vartheta^2}\right]_{\vartheta_v}}$$

cantidad de información esperada en la muestra respecto al parámetro ϑ. La cota de Cramer-Rao puede escribirse:

$$\text{Var}(\hat{\vartheta}_c) \geq IE(\vartheta)^{-1}$$

alternativamente:

$$\boxed{\text{Eficiencia}\,(\hat{\vartheta}_c) \leq IE(\vartheta)}$$

que nos dice que la eficiencia o precisión de cualquier estimador centrado es menor o igual que la cantidad de información esperada en la muestra. Cuando coinciden, la precisión es máxima, e igual a la cantidad de información esperada.

La cantidad de información es aditiva: si llamamos $ie(\vartheta)$ a la cantidad de información en una muestra de tamaño 1, se verifica que en una muestra aleatoria simple:

Cuadro 7.3 La función soporte, la tasa de discriminación y cantidad de información observada para algunos modelos regulares

Modelo	1: Binomial	2: Poisson
ϑ	p	λ
$\ell(\vartheta)$	$\vartheta^r(1-\vartheta)^{n-r}$	$e^{-n\vartheta}\vartheta^{n\bar{x}}$
$L(\vartheta)$	$n\left(\ln[1-\vartheta]+\frac{r}{n}\ln\frac{\vartheta}{1-\vartheta}\right)$	$n(\bar{x}\ln\vartheta-\vartheta)$
$\frac{dL(\vartheta)}{d\vartheta}$	$\frac{r}{\vartheta}+\frac{(r-n)}{1-\vartheta}$	$n\left(\frac{\bar{x}}{\vartheta}-1\right)$
$E\left[\frac{dL(\vartheta)}{d\vartheta}\right]_{\vartheta_v}$	$\frac{n\vartheta_v}{\vartheta_v}+\frac{n\vartheta_v-n}{1-\vartheta_v}=0$	$n\left(\frac{\vartheta_v}{\vartheta_v}-1\right)=0$
$\frac{d^2L(\vartheta)}{d\vartheta^2}$	$-\frac{r}{\vartheta^2}+\frac{(r-n)}{(1-\vartheta)^2}$	$-n\frac{\bar{x}}{\vartheta^2}$
$E\left[-\frac{d^2L(\vartheta)}{d\vartheta^2}\right]_{\vartheta_v}$	$\frac{n}{\vartheta_v(1-\vartheta_v)}$	$\frac{n}{\vartheta_v}$

Nota: ϑ_v representa el verdadero valor del parámetro, $s^2=\frac{\sum(x_i-\mu)^2}{n}$, $\text{Var}(\bar{x})=\frac{\sigma^2}{n}$, $E(s^2)=\sigma^2$.

$$IE(\vartheta) = n\ ie(\vartheta)$$

Análogamente, al término:

$$\boxed{IO(\hat{\vartheta}_{MV}) = -\frac{d^2L(\hat{\vartheta}_{MV})}{d\vartheta^2}}$$

se le denomina *cantidad de información observada* en la muestra. Observemos que este término no depende de ϑ —a diferencia de $IE(\vartheta)$—, ya que se calcula sustituyendo el parámetro ϑ por su estimación máximo-verosímil, $\hat{\vartheta}_{MV}$. Se demuestra que asintóticamente la información observada coincide con la esperada, resultado que se ha utilizado al escribir la varianza asintótica del estimador *MV* en (7.24).

3: Exponencial	4: Normal, σ conocida	5: Normal, μ conocida
λ	μ	σ^2
$\vartheta^n e^{-\vartheta n\bar{x}}$	$e^{-n(\bar{x}-\vartheta)^2/2\sigma^2}$	$\vartheta^{-n/2} e^{-ns^2/2\sigma}$
$n(\ln \vartheta - \vartheta\bar{x})$	$-\frac{n}{2\sigma^2}(\bar{x}-\vartheta)^2$	$-\frac{n}{2}\left(\ln \vartheta + \frac{s^2}{2\vartheta}\right)$
$n\left(\frac{1}{\vartheta} - \bar{x}\right)$	$\frac{n}{\sigma^2}(\bar{x}-\vartheta)$	$\frac{n}{2\vartheta}\left(\frac{s^2}{\vartheta} - 1\right)$
$n\left(\frac{1}{\vartheta_v} - \frac{1}{\vartheta_v}\right) = 0$	$\frac{n}{\sigma^2}(\vartheta_v - \vartheta_v) = 0$	$\frac{n}{2\vartheta_v}\left(\frac{\vartheta_v}{\vartheta_v} - 1\right) = 0$
$-\frac{n}{\vartheta^2}$	$-\frac{n}{\sigma^2}$	$-\frac{n}{2}\left(\frac{2s^2}{\vartheta^3} - \frac{1}{\vartheta^2}\right)$
$\frac{n}{\vartheta_v^2}$	$\frac{n}{\sigma^2}$	$\frac{n}{2\vartheta_v^2}$

Estos resultados se generalizan para un parámetro vectorial $\hat{\boldsymbol{\vartheta}}$. Si $\hat{\boldsymbol{\vartheta}}_c$ es un estimador centrado

$$E(\hat{\boldsymbol{\vartheta}}_c) = \boldsymbol{\vartheta}$$

llamando $\mathbf{Var}(\hat{\boldsymbol{\vartheta}}_c)$ a la matriz de varianzas y covarianzas del estimador y $\mathbf{H}(\vartheta)$ a la matriz hessiano de segundas derivadas y $\mathbf{IE}(\boldsymbol{\vartheta})$ a la *matriz de información esperada:*

$$\boxed{\mathbf{IE}(\boldsymbol{\vartheta}) = E[-\mathbf{H}(\boldsymbol{\vartheta})] = E\left[-\frac{\partial^2 L(\boldsymbol{\vartheta})}{\partial\vartheta_i \partial\vartheta_j}\right]}$$

se demuestra que la matriz de varianzas y covarianzas de los estimadores, $\mathbf{Var}(\hat{\boldsymbol{\vartheta}}_c)$, es «mayor» que $\mathbf{IE}(\boldsymbol{\vartheta})^{-1}$ en el sentido de que:

$$\mathbf{Var}(\hat{\boldsymbol{\vartheta}}_c) - \mathbf{IE}(\boldsymbol{\vartheta})^{-1} \text{ es semidefinida positiva}$$

En el caso vectorial, la *matriz de información observada* se define, siendo $\hat{\boldsymbol{\vartheta}}_{MV}$ el estimador máximo verosímil, por:

$$\mathbf{IO}(\hat{\boldsymbol{\vartheta}}_{MV}) = \mathbf{H}(\hat{\boldsymbol{\vartheta}}_{MV}) = \left[-\frac{\partial^2 L(\hat{\boldsymbol{\vartheta}}_{MV})}{\partial \vartheta_i \partial \vartheta_j} \right]$$

y coincide asintóticamente con la matriz de información esperada. Para familias regulares el estimador *MV* es eficiente y podemos calcular su varianza asintótica con la información observada.

8. Estimación por intervalos

Jerzy Neyman (1894-1981)
Científico ruso de origen polaco. Creador, con E. Pearson, de la teoría estadística de contraste de hipótesis, de la teoría de investigación por muestreo y de la estimación por intervalos de confianza. Emigró a Londres y después a Estados Unidos, donde fundó el Departamento de Estadística de la Universidad de California en Berkeley.

8.1 Introducción

En el capítulo anterior hemos visto cómo obtener estimadores para un parámetro y cómo calcular una medida de la precisión del estimador: su desviación típica en el muestreo. Proporcionar un estimador sin indicar su precisión es de escasa utilidad y puede ser engañoso. Por esta razón siempre conviene dar junto al estimador un intervalo de valores entre los cuales deberá estar el valor del parámetro de interés con alta probabilidad. Éste es el objetivo de la estimación por intervalos.

Para ilustrar el método de construcción de los intervalos de confianza, consideremos como ejemplo la estimación de la media con una muestra de tamaño 25 en una población normal de desviación típica conocida e igual a 10. Antes de observar la muestra y calcular el estimador, $\bar{x}$, podemos hacer predicciones de las discrepancias esperadas entre el estimador y el parámetro. Por ejemplo, podemos prever que el 95% de las veces:

$$|\bar{x} - \mu| \leq 1{,}96 \frac{\sigma}{\sqrt{n}} = 1{,}96 \cdot \frac{10}{\sqrt{25}} = 3{,}92$$

es decir, el 95% de las veces $\bar{x}$ no será μ más de 3,92 unidades. En consecuencia, si observamos $\bar{x} = 40$, podemos concluir que μ estará previsiblemente en el intervalo $40 \pm 3{,}92$. Ésta es la idea central de construcción de un intervalo de confianza, que vamos a analizar con detalle.

Llamaremos intervalo de confianza para el parámetro ϑ con nivel o coeficiente de confianza $1 - \alpha$, a una expresión del tipo:

$$\vartheta_1 \leq \vartheta \leq \vartheta_2$$

donde los límites ϑ_1 y ϑ_2 dependen de la muestra y se calculan de manera tal que si tomamos muchas muestras, todas del mismo tamaño, y construimos un intervalo con cada una, podemos afirmar que el $100 \cdot (1 - \alpha)\%$ de los intervalos así construidos contendrán el verdadero valor del parámetro. Por ejemplo, el intervalo $\bar{x} \pm 3{,}92$ anterior tiene la propiedad de que si tomamos muchas muestras de tamaño 25 de esa población normal y construimos con cada muestra un intervalo de confianza (los intervalos serán distintos, porque $\bar{x}$ variará de muestra en muestra), el 95% de los intervalos así construidos contendrán el verdadero valor de la media. La razón es simple: contendrán la media siempre que $|\bar{x} - \mu| \leq 3{,}92$, y, por construcción, esto ocurrirá con probabilidad 0,95.

Observemos que la clave del procedimiento anterior es que el error relativo de estimación, dado por

$$\frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

si σ es conocida, tiene una distribución totalmente conocida (normal estándar) que no depende del valor desconocido de μ. Por tanto, si fijamos una probabilidad podemos encontrar un intervalo de valores entre los que estará este error relativo con esa probabilidad y luego despejar el parámetro desconocido para obtener su intervalo. Generalizando esta idea podemos concluir que, si fijamos α, el problema de determinar ϑ_1 y ϑ_2 puede resolverse cuando exista una variable aleatoria, definida como una función de ϑ y de los datos muestrales, cuya distribución está perfectamente determinada y sea la misma para cualquier valor del parámetro. Para muestras grandes esta variable siempre existe, ya que si estimamos ϑ por su estimador máximo-verosímil, $\hat{\vartheta}_{MV}$, el error relativo de estimación definido por:

$$\omega = \frac{\vartheta - \hat{\vartheta}_{MV}}{\sigma(\hat{\vartheta}_{MV})} \tag{8.1}$$

donde $\sigma(\hat{\vartheta}_{MV})$ es la desviación típica asintótica de la distribución muestral del estadístico máximo-verosímil que sigue, asintóticamente, una distribución normal estándar. Como $\sigma(\hat{\vartheta}_{MV})$ es conocido, ω es sólo función de ϑ y tiene una distribución totalmente conocida.

Este resultado es fundamental: *indica que, sea cual sea ϑ, podemos conocer aproximadamente la distribución del error relativo que cometeremos al estimar este parámetro por $\hat{\vartheta}_{MV}$.*

En el caso general, sea $\omega = g(\vartheta; \mathbf{X})$ una variable con distribución conocida, que admitiremos es función continua y monótona de ϑ, y que llamaremos *estadístico pivote* para el intervalo. Entonces, dado α, es posible encontrar valores a y b tales que

$$P(a \leq g[\vartheta, \mathbf{X}] \leq b) = 1 - \alpha \tag{8.2}$$

Por la hipótesis de que g es una función continua y monótona de ϑ, la expresión anterior equivale a:

$$P(g^{-1}[a, \mathbf{X}] \leq \vartheta \leq g^{-1}[b, \mathbf{X}]) = 1 - \alpha \tag{8.3}$$

y el intervalo de nivel $1 - \alpha$ para ϑ será, llamando $\vartheta_1 = g^{-1}(a, \mathbf{X})$ y $\vartheta_2 = g^{-1}(b, \mathbf{X})$:

$$\vartheta_1 \leq \vartheta \leq \vartheta_2$$

Observemos que, por el procedimiento seguido, el intervalo anterior contendrá el verdadero valor del parámetro siempre que ω esté contenido entre a y b, lo que ocurrirá, por construcción, el $100(1 - \alpha)\%$ de las veces.

El método anterior plantea tres interrogantes: el primero, cómo encontrar el estadístico pivote $\omega = g(\vartheta, \mathbf{X})$; el segundo, cómo seleccionar los valores a y b en la distribución de $g(\vartheta, \mathbf{X})$; el tercero, cómo elegir un valor α para construir el intervalo. Vamos a analizar estos tres aspectos.

8.2 Metodología

8.2.1 La selección del estadístico pivote

Definamos, el error relativo de estimación (ω) por:

$$\text{error relativo } (\omega) = \frac{\text{error cometido}}{\text{error promedio}}$$

consideremos los casos siguientes:

a) Cuando ϑ es un parámetro de tendencia central, se verifica:

$$\omega = \frac{\vartheta - \hat{\vartheta}_{MV}}{\hat{\sigma}} \tag{8.4}$$

donde $\hat{\sigma}$ es una estimación de la dispersión. Este error es adimensional y será, por tanto, invariante ante cambios de escala.

Tomando $\hat{\sigma}^2 = \widehat{\text{Var}}(\hat{\vartheta}_{MV})$, la distribución asintótica de ω es $N(0, 1)$. Para pequeñas muestras es posible en muchos casos seleccionar otro valor de $\hat{\sigma}$ que conduzca a una distribución *t* de Student, como veremos en las secciones siguientes.

b) Cuando ϑ sea un parámetro de variabilidad (varianza, Meda, etc.), el error relativo será:

$$\frac{\vartheta - \hat{\vartheta}_{MV}}{\vartheta} = 1 - \frac{\hat{\vartheta}_{MV}}{\vartheta} \tag{8.5}$$

que es de nuevo adimensional. El cociente $\hat{\vartheta}_{MV}/\vartheta$ sigue, en poblaciones normales, una distribución conocida que nos proporcionará el intervalo.

c) Para parámetros generales ϑ, la variable ω se construye buscando funciones $g(\vartheta, \hat{\vartheta}_{MV}[\mathbf{X}])$ que sean adimensionales y tengan una distribución simple que no dependa de ϑ y pueda tabularse.

La razón de buscar funciones $g(\vartheta, \mathbf{X})$ basándonos en el estadístico máximo-verosímil es conseguir intervalos lo más cortos posible. La distancia (a, b) depende de la varianza de la variable $\omega = g(\vartheta, \mathbf{X})$ y, por tanto, es deseable partir de estimadores con varianza lo más pequeña posible, lo que nos conduce a los estimadores máximo-verosímiles.

8.2.2 La determinación de los límites

Una vez seleccionado el estadístico pivote, queda el problema de determinar los límites *a* y *b*. Un criterio razonable es escoger estos valores de manera que el intervalo sea de longitud mínima. Si la distribución de $\omega = g(\vartheta, \mathbf{X})$ es simétrica y unimodal, esto se consigue tomando el intervalo centrado alrededor del valor central, dejando $\alpha/2$ de probabilidad a ambos lados. Si la distribución de ω es asimétrica, la determinación de los valores (a, b) es complicada y, por simplicidad, los tomaremos simétricamente, como en el caso anterior.

Distribución de confianza

El tercer interrogante, cómo escoger el valor de α, se resuelve habitualmente tomando un valor arbitrario pequeño, como 0,05 o 0,01. Un procedimiento más informativo es presentar la distribución de probabilidad que va a generar todos los intervalos posibles para cada valor de α. A esta distribución la llamaremos *distribución de confianza*, y resulta de considerar los datos muestrales como fijos y el parámetro como una variable aleatoria. Por ejemplo, en (8.1), la distribución de confianza de ϑ se obtendrá despejando ϑ:

$$\vartheta = \hat{\vartheta}_{MV} + \omega\, \sigma(\hat{\vartheta}_{MV})$$

como ω es $N(0, 1)$, ϑ será normal con media $\hat{\vartheta}_{MV}$ y desviación típica $\sigma(\hat{\vartheta}_{MV})$. En general, despejaremos ϑ para obtener:

$$\vartheta = g^{-1}(\omega, \mathbf{X}) \tag{8.6}$$

y la distribución de ϑ así obtenida es la distribución de confianza.

Su nombre proviene de que, llamando $f(\vartheta|\mathbf{X})$ a esta distribución:

$$\int_{\vartheta_1}^{\vartheta_2} f(\vartheta \mid \mathbf{X})\, d\vartheta = 1 - \alpha \Rightarrow \int_{\omega(\vartheta_1)}^{\omega(\vartheta_2)} f(\omega) d\omega = 1 - \alpha$$

aplicando la fórmula del cambio de variable y teniendo en cuenta que, por hipótesis, la relación entre ω y ϑ es biunívoca. Entonces el intervalo $\omega(\vartheta_2) \leq \omega \leq \omega(\vartheta_1)$ equivale al $\vartheta_2 \leq \vartheta \leq \vartheta_1$

Aunque esta notación es conveniente y útil, esta distribución no puede interpretarse estrictamente como la distribución de probabilidad de ϑ, ya que al ser éste un valor fijo, aunque desconocido, no le asignaremos probabilidades (compárese con el enfoque bayesiano del capítulo siguiente).

8.3 Intervalos para medias de poblaciones normales

8.3.1 Varianza conocida

Si conocemos la varianza de la población, sabemos que el error relativo de estimación de μ mediante la media muestral $\bar{x}$:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \tag{8.7}$$

es una variable normal estándar. Por lo tanto, esta variable, z, permite construir un intervalo de confianza. Tendremos:

$$P\left(-z_{\alpha/2} \leq \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) = 1-\alpha$$

donde $z_{\alpha/2}$ es un valor de la normal estándar tal que:

$$P(z > z_{\alpha/2}) = 1 - \Phi(z_{\alpha/2}) = \alpha/2$$

siendo Φ la función de distribución normal estándar. Entonces, el intervalo será:

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{8.8}$$

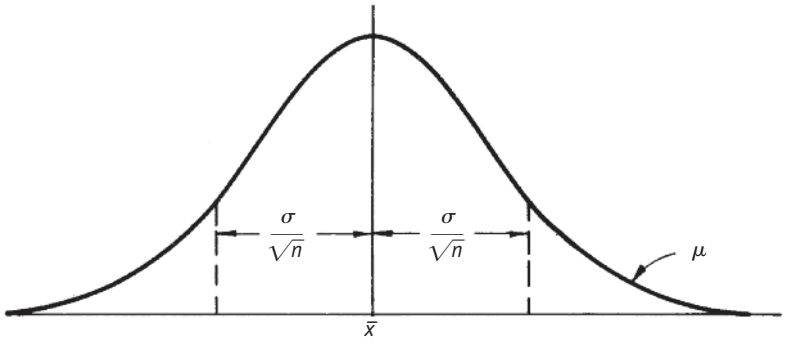
y tendrá de confianza $(1-\alpha)$. Los valores se han escogido simétricos para que conduzcan al intervalo más corto posible.

La distribución de confianza en este caso resulta al despejar μ en (8.7) para obtener:

$$\mu = \bar{x} + z\frac{\sigma}{\sqrt{n}}$$

Al variar α, como $\bar{x}$ se supone constante, la única variable aleatoria es z; por tanto, la distribución generada es normal, con media $\bar{x}$ y varianza $\sigma/\sqrt{n}$. Esta distribución resume la incertidumbre existente respecto al valor desconocido μ (véase la figura 8.1).

Figura 8.1 Distribución de confianza para la media de una población normal, σ conocida.



8.3.2 Varianza desconocida

Si σ es desconocida, no podemos utilizar la expresión anterior y acudiremos a la distribución t. Observemos que:

$$t = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} : \sqrt{\frac{(n-1)\hat{s}^2}{\sigma^2(n-1)}} = \frac{\bar{x} - \mu}{\hat{s}/\sqrt{n}} \tag{8.9}$$

es el cociente entre una variable $N(0, 1)$ y la raíz de una distribución χ^2_g/g, siendo además numerador y denominador independientes. Por tanto, esta variable sigue una distribución t con $n - 1$ grados de libertad. Además, el estadístico obtenido no depende de σ, y es función monótona de μ. Entonces, si $P(t > t_{\alpha/2}) = \alpha/2$, el intervalo será:

$$\bar{x} - t_{\alpha/2}\frac{\hat{s}}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2}\frac{\hat{s}}{\sqrt{n}} \tag{8.10}$$

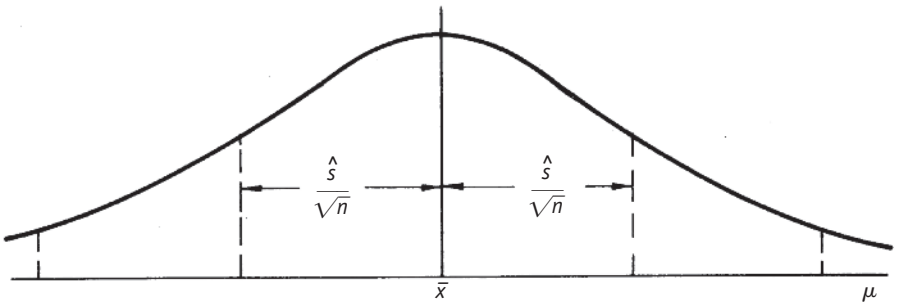
y tendrá confianza $1 - \alpha$.

La distribución de confianza será ahora la definida por:

$$\mu = \bar{x} + t\frac{\hat{s}}{\sqrt{n}}$$

Si suponemos $\bar{x}$, $\hat{s}$ y n fijos, la distribución inducida por t es la t generalizada con media $\bar{x}$ y factor de escala $\hat{s}/\sqrt{n}$, como indica la figura 8.2.

Figura 8.2 Distribución de confianza para la media, σ desconocida



Ejemplo 8.1

El director de una empresa ha anunciado que los salarios el año pasado crecieron un promedio del 3,5%. Un grupo de trabajadoras toma una muestra de los incrementos que han recibido una muestra de 10 mujeres obteniendo los siguientes incrementos: 3%, 3%, 5%, 1%, 1%, 2%, 1%, 1,5%, 2%, 2%. Construir un intervalo de confianza para el incremento medio experimentado por la remuneración de las mujeres en esta empresa.

La media de los incrementos es

$$\bar{x} = \frac{3 + 3 + 5 + 1 + 1 + 2 + 1 + 1 + 5 + 2 + 2}{10} = 2{,}36$$

y la desviación típica

$$\hat{s} = \sqrt{\frac{(3 - 2{,}36)^2 + \ldots + (2 - 2{,}36)^2}{9}} = 1{,}50$$

El intervalo de confianza del 95% requiere el percentil de la distribución t de Student con 9 grados de libertad que es 2,26. En consecuencia el intervalo será:

$$2{,}36 - 2{,}26 \frac{1{,}5}{\sqrt{10}} \leq \mu \leq 2{,}36 + 2{,}26 \frac{1{,}5}{\sqrt{10}}$$

que resulta en el intervalo (1,29%-3,44%). Este intervalo no incluye el 3,5% como valor posible, por lo que podemos concluir que existe una fuerte evidencia de que las mujeres han recibido un incremento salarial menor que la media de los trabajadores.

8.4 Intervalos para medias. Caso general

Para cualquier población la media muestral es asintóticamente normal con media μ y desviación típica $\sigma/\sqrt{n}$. Por tanto, para muestras grandes de cualquier población, el intervalo de confianza para la media es:

$$\bar{x} - z_{\alpha/2} \frac{\hat{s}}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \frac{\hat{s}}{\sqrt{n}} \tag{8.11}$$

donde se ha utilizado $\hat{s}$ como estimador de σ.

8.4.1 Proporciones

Cuando se desea estimar la proporción (p) de elementos con un atributo, la población base es de Bernoulli y la media muestral es el cociente el número de elementos con el atributo estudiado (r) *y* el tamaño muestral (n). La varianza muestral será:

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n} = \frac{1}{n}\,[r(1-\hat{p})^2 + (n-r)(0-\hat{p})^2] =$$

$$\frac{1}{n}\left(\frac{r[n-r]^2}{n^2} + \frac{[n-r]r^2}{n^2}\right) = \hat{p}\hat{q}$$

y la varianza de la distribución muestral de la media, estimada por s^2/n, se convierte en $\hat{p}\hat{q}/n$ (podría utilizarse $\hat{s}$ en lugar de s, pero como suponemos tamaño muestral grande la diferencia es irrelevante). Entonces el intervalo será:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} \qquad (8.12)$$

y es un caso particular del método general anterior. En el ejemplo siguiente se obtiene este intervalo utilizando las propiedades de los estimadores *MV*.

8.5 Intervalo para varianzas de poblaciones normales

Para construir un intervalo para la varianza de una población normal, tenemos en cuenta que:

$$\frac{ns^2}{\sigma^2} = \frac{(n-1)\hat{s}^2}{\sigma^2} \qquad (8.13)$$

se distribuye como una χ^2_{n-1}. Por lo tanto, determinando dos valores χ^2_a y χ^2_b que dejen entre sí el $1 - \alpha$ de la distribución:

$$P\left(\chi^2_a \le \frac{ns^2}{\sigma^2} \le \chi^2_b\right) = 1 - \alpha$$

$$P\left(\frac{1}{\chi^2_a} \ge \frac{\sigma^2}{ns^2} \ge \frac{1}{\chi^2_b}\right) = 1 - \alpha$$

obtenemos el intervalo:

$$\boxed{\frac{ns^2}{\chi^2_a} \geq \sigma^2 \geq \frac{ns^2}{\chi^2_b}} \tag{8.14}$$

La distribución de confianza será ahora, despejando el parámetro σ^2:

$$\sigma^2 = \frac{ns^2}{\chi^2}$$

que se conoce como *distribución* χ^2 *invertida* y es asimétrica.

Ejemplo 8.2

Se han medido los siguientes valores (en miles de personas) para la audiencia de un programa de televisión en distintos días: 521, 742, 593, 635, 788, 717, 606, 639, 666, 624. Construir un intervalo de confianza para la audiencia media y otro para la varianza, en la hipótesis de normalidad.

La estimación de la media será:

$$\Sigma\, x_i = 521 + 742 + \ldots + 624 = 6531$$

$$\bar{x} = \frac{6531}{10} = 653{,}1$$

y la de la varianza:

$$\Sigma\, x_i^2 = 521^2 + 742^2 + \ldots + 624^2 = 4320401$$

$$s^2 = \frac{\Sigma x_i^2}{10} - \bar{x}^2 = \frac{4320401}{10} - 653{,}1^2 = 5500{,}49$$

$$\hat{s}^2 = \frac{10}{9}\, s^2 = 6111{,}66$$

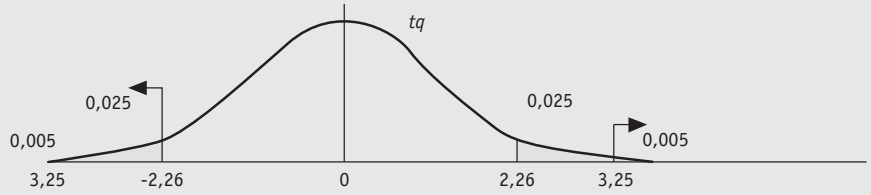
Por tanto, el intervalo para la media será:

$$\mu \in 653{,}1 \pm t_{\alpha/2} \sqrt{\frac{6111{,}66}{10}}$$

$$\mu \in 653{,}1 \pm t_{\alpha/2}\ 24{,}72$$

Si fijamos $\alpha = 0{,}05$, como la t tiene $n - 1 = 10 - 1 = 9$ grados de libertad, en tablas se obtienen los valores ± 2,26 para el 95% y ± 3,25 para el 99% (véase la figura 8.3).

Figura 8.3



Y, por tanto, el intervalo del 95% será:

$$\mu \in 653{,}1 \pm 2{,}26 \cdot 24{,}72$$

es decir, la audiencia medida en miles de persona está en el intervalo:

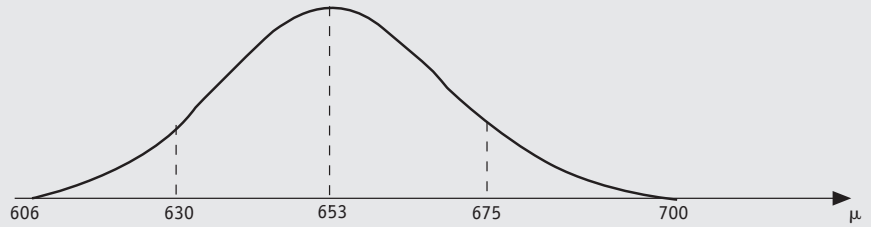
$$(597{,}23 \quad ; \quad 708{,}96)$$

la distribución de confianza para μ, de la que obtenemos cualquier intervalo, será:

$$\mu = 653{,}1 + t\ 24{,}72$$

y tendrá la forma de la figura 8.4.

Figura 8.4



Para construir un intervalo de confianza para σ^2, como

$$\frac{(n-1)\hat{s}^2}{\sigma^2} = \frac{9 \cdot 6111{,}66}{\sigma^2} \text{ es } \chi^2_9$$

y en tablas se obtienen, con $\alpha = 0{,}05$, y tomando el intervalo simétrico, los valores 2,7 y 19 para la χ^2_9. Tendremos:

$$\frac{9 \cdot 6111{,}66}{2{,}7} \geq \sigma^2 \geq \frac{9 \cdot 6111{,}66}{19}$$

$$20{,}372 \geq \sigma^2 \geq 2895$$

y un intervalo aproximado para σ, tomando raíces

$$142{,}73 \geq \sigma \geq 53{,}81$$

En resumen, los datos indican que la mejor estimación de media de la distribución es 653 y es muy improbable que el verdadero valor esté fuera del intervalo (597, 709). La mejor estimación de la desviación típica es 75 ($\sqrt{5500{,}49}$) y es improbable que el verdadero valor esté fuera del intervalo (54, 143) para σ.

8.6 Intervalo para la diferencia de medias, poblaciones normales

8.6.1 Caso de varianzas iguales

Supongamos dos poblaciones normales $N(\mu_1, \sigma)$, $N(\mu_2, \sigma)$ con la misma varianza. Tenemos dos muestras independientes $(x_{11}, ..., x_{1n1})$ y $(x_{21}, ..., x_{2n2})$ de ambas poblaciones y queremos hacer un intervalo de confianza para la diferencia de medias. Llamando $\bar{x}_1, \bar{x}_2$ a las medias y $\hat{s}_1^2, \hat{s}_2^2$ a las varianzas corregidas, tendremos que:

$$\frac{\bar{x}_1 - \mu_1}{\sigma/\sqrt{n_1}} \quad \text{y} \quad \frac{\bar{x}_2 - \mu_2}{\sigma/\sqrt{n_2}} \text{ son } N(0, 1)$$

$$\frac{(n_1 - 1)\hat{s}_1^2}{\sigma^2} \quad \text{y} \quad \frac{(n_2 - 1)\hat{s}_2^2}{\sigma^2} \text{ son } \chi^2 \text{ con } n_1 - 1 \text{ y } n_2 - 1 \text{ g. de l.}$$

Llamando

$$\hat{s}_T^2 = \frac{(n_1 - 1)\hat{s}_1^2 + (n_2 - 1)\hat{s}_2^2}{n_1 + n_2 - 2} \tag{8.15}$$

a la estimación de la varianza común que es, según lo estudiado en la sección 4, una media ponderada de las estimaciones independientes $\hat{s}_1^2$ y $\hat{s}_2^2$ con pesos de ponderación sus precisiones que, en el caso de varianzas de poblaciones normales, son proporcionales a sus grados de libertad. Como la suma de variables χ^2 independientes es otra distribución χ^2 con grados de libertad la suma de los de ambas:

$$\frac{(n_1+n_2-2)\hat{s}_T^2}{\sigma^2}=\frac{(n_1-1)\hat{s}_1^2}{\sigma^2}+\frac{(n_2-1)\hat{s}_2^2}{\sigma^2} \text{ es } \chi^2_{n_1+n_2-2}$$

por tanto, podemos construir una distribución t partiendo de la variable $z=\bar{x}_1-\bar{x}_2$ que tendrá media $\mu_1-\mu_2$ y varianza:

$$\text{Var}(z)=\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)$$

La variable

$$t=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sigma\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)^{1/2}} \;:\; \frac{\hat{s}_T}{\sigma}=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\hat{s}_T\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)^{1/2}} \qquad (8.16)$$

es una distribución t con grados de libertad n_1+n_2-2. La distribución de confianza para la diferencia de medias será:

$$(\mu_1-\mu_2)=\bar{x}_1-\bar{x}_2+t_{n_1+n_2-2}\hat{s}_T\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{1/2}$$

que es una distribución t general, con parámetros $\bar{x}_1-\bar{x}_2$ y $\hat{s}_T\sqrt{1/n_1+1/n_2}$. El intervalo de confianza de nivel $1-\alpha$ será:

$$\boxed{\mu_1-\mu_2\in\left(\bar{x}_1-\bar{x}_2\pm t_{\alpha/2(n_1+n_2-2)}\hat{s}_T\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)} \qquad (8.17)$$

donde entre paréntesis en la t aparecen sus grados de libertad.

8.6.2 Caso de varianzas desiguales

Cuando las varianzas de ambas poblaciones no pueden suponerse iguales, se utiliza el siguiente procedimiento aproximado: un intervalo al nivel $1-\alpha$ es:

$$\mu_1 - \mu_2 \in \left(\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2[g]} \sqrt{\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2}} \right)$$

donde g, grados de libertad de la t, es $n_1 + n_2 - 2 - \Delta$, siendo Δ un número positivo corrector que se calcula tomando el entero más próximo a:

$$\Delta = \frac{[(n_2 - 1)S_1 - (n_1 - 1)S_2]^2}{(n_2 - 1)S_1^2 - (n_1 - 1)S_2^2} \tag{8.18}$$

siendo $S_i = \hat{s}_1^2/n_i$ ($i = 1,2$). Se comprueba que:

$$0 \leq \Delta \leq \text{máx}(n_1 - 1, n_2 - 1)$$

La interpretación del término corrector (8.18) es la siguiente: si la primera población tiene mucha mayor varianza que la segunda y $n_1 = n_2$, entonces $s_1^2 >> s_2^2$ y $\Delta \simeq n_2 - 1$ con lo que $g = n_1 - 1$ y los grados de libertad dependen de la precisión con que estimemos la varianza de la primera población. Si las varianzas de ambas poblaciones son similares y también los tamaños muestrales, el término corrector se anula y estamos en el caso anterior. Finalmente, si los tamaños muestrales son muy distintos y, por ejemplo, $n_1 >> n_2$, el término corrector será alto y los grados de libertad de la t se reducen.

8.7 Diferencias de medias. Caso general

Para tamaños muestrales grandes la variable $y = \bar{x}_1 - \bar{x}_2$ será asintóticamente normal con media $\mu_1 - \mu_2$ y varianza la suma de varianzas. Por tanto, un intervalo aproximado para muestras grandes es:

$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2} \sqrt{\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2} \sqrt{\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2}} \tag{8.19}$$

Proporciones

Como paso particular del resultado anterior, si la población es de Bernoulli, la media muestral es la proporción observada y el intervalo será:

$$(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2} \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} \leq p_1 - p_2 \leq (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2} \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} \tag{8.20}$$

8.8 Intervalo para la razón de varianzas en poblaciones normales

Dadas dos muestras independientes de tamaños n_1 y n_2 de dos poblaciones normales, se verifica que, $\hat{s}_i^2/\sigma_i^2$ sigue una distribución χ^2 $(n_i - 1)$ dividida por sus grados de libertad. El cociente entre las dos poblaciones:

$$F(_{n_1-1,\, n_2-1}) = \frac{\hat{s}_1^2/\sigma_1^2}{\hat{s}_2^2/\sigma_2^2} = \frac{\hat{s}_1^2}{\hat{s}_2^2}\frac{\sigma_2^2}{\sigma_1^2}$$

seguirá, por construcción, una distribución F con $(n_1 - 1)$ y $(n_2 - 1)$ grados de libertad. En consecuencia, construiremos un intervalo de nivel α para la razón de varianzas utilizando que:

$$F_a \frac{\hat{s}_2^2}{\hat{s}_1^2} \leq \frac{\sigma_2^2}{\sigma_1^2} \leq F_b \frac{\hat{s}_2^2}{\hat{s}_1^2} \tag{8.21}$$

donde los valores F_a y F_b se determinan en las tablas de la distribución F con la condición $P(F_a \leq F \leq F_b) = 1 - \alpha$.

Ejemplo 8.3

El número diario de clientes atendidos en un puesto de servicio *A* en cinco días ha sido: 50, 48, 53, 60, 37; mientras que, en esos mismos días, un puesto *B* ha atendido: 40, 51, 62, 55 y 64. Se pide:

a) Construir un intervalo para la diferencia de demanda media entre los puestos de servicio *A* y *B*, suponiendo la misma desviación típica.
b) Lo mismo, pero sin suponer la misma desviación.
c) Determinar cuál debía haber sido el tamaño muestral *n* de ambas muestras para que en el caso (a) y con el mismo valor de la varianza estimada la longitud del intervalo para la diferencia de medias con $\alpha = 0{,}05$ fuese de 8 unidades.

a)
$$\bar{x}_1 = \frac{50 + 48 + 53 + 60 + 37}{5} = 49{,}6$$

$$\hat{s}_1 = \sqrt{\frac{(50 - 49{,}6)^2 + \ldots + (37 - 49{,}6)^2}{4}} = 8{,}38$$

$$\bar{x}_2 = \frac{40 + 51 + \ldots + 64}{5} = 54{,}4$$

$$\hat{s}_2 = \sqrt{\frac{(40 - 54{,}4)^2 + \ldots + (64 - 54{,}4)^2}{4}} = 9{,}61$$

Entonces, como:

$$\bar{x}_1 - \bar{x}_2 = 49{,}6 - 54{,}4 = -4{,}80$$

$$\hat{s}_T^2 = \frac{8{,}38^2 + 9{,}61^2}{2} = 81{,}29; \qquad \hat{s}_T = 9{,}02$$

el intervalo será:

$$\mu_1 - \mu_2 \in \left(-4{,}80 \pm t_{(8)} \cdot 9{,}02 \sqrt{\frac{2}{5}} \right)$$

donde $t(8)$ indica que es un valor de t con 8 grados de libertad. Tomando $\alpha = 0{,}05$, $t_{0{,}975}(8) = 2{,}31$ y tendremos:

$$\mu_1 - \mu_2 \in (-4{,}80 \pm 2{,}31 \cdot 9{,}02 \sqrt{2/5})$$

es decir:

$$\mu_1 - \mu_2 \in (-4{,}80 \pm 13{,}18)$$

que indica que el intervalo para la diferencia de medias es (–17,98; 8,38) y la diferencia verdadera puede ser mayor o menor que cero.

b) Eliminando la hipótesis de igualdad de varianzas, el intervalo será:

$$-4{,}80 \pm t_{(\alpha/2;\, g)} \sqrt{8{,}38^2/5 + 9{,}61^2/5}$$

llamando:

$$S_1 = 8{,}38^2/5 = 14{,}04$$
$$S_2 = 9{,}61^2/5 = 18{,}47$$

el término corrector de los grados de libertad es:

$$\Delta = \frac{(4 \cdot 14{,}04 - 4 \cdot 18{,}47)^2}{4 \cdot (14{,}04^2 + 18{,}47^2)} = \frac{314}{2\,153} = 0{,}15$$

Por tanto, tomaremos $\Delta = 0$ y la t tendrá 8 grados de libertad como en el caso (a). Con $\alpha = 0{,}05$ el intervalo será:

$$-4{,}80 \pm 2{,}31 \cdot 5{,}7 = -4{,}80 \pm 13{,}17$$

que es casi idéntico al anterior. Construiremos un intervalo para el cociente de varianzas. Tomando $F_a = F_{1-\alpha/2}$ y $\alpha = 0{,}10$, según la tabla 7 del apéndice:

$$F(4{,}4;\ 0{,}95) = F^{-1}(4{,}4;\ 0{,}05) = (6{,}3883)^{-1}$$

y el intervalo resulta:

$$(9{,}61/8{,}38)^2(6{,}3883)^{-1} \leq \sigma_2^2/\sigma_1^2 \leq (9{,}61/8{,}38)^2(6{,}3883)$$

$$0{,}21 \leq \sigma_2^2/\sigma_1^2 \leq 8{,}40$$

lo que sugiere que es perfectamente posible que las varianzas sean iguales ($\sigma_2^2/\sigma_1^2 = 1$), ya que este punto está incluido en el intervalo de confianza.

c) Si queremos que la longitud del intervalo sea 8, la semilongitud es:

$$4 = t(8) \cdot 9{,}02\sqrt{\frac{1}{n} + \frac{1}{n}}$$

despejando n

$$n = \frac{2 \cdot t(8)^2 \cdot 9{,}02^2}{16}$$

y sustituyendo

$$n = \frac{2 \cdot 2{,}31^2 \cdot 9{,}02^2}{16} = \frac{868}{16} = 54{,}2$$

Por tanto, si hubiésemos tomado $n \geq 55$ días en cada puesto de servicio, y suponiendo que la desviación típica estimada sería próxima al valor encontrado, 9,02, tendríamos un intervalo del 95% de longitud total 4 unidades.

8.9 Intervalos asintóticos

Si ϑ es cualquier parámetro de una población y $\hat{\vartheta}_{MV}$ su estimación máximo-verosímil, sabemos que, asintóticamente:

$$E(\hat{\vartheta}_{MV}) \to \vartheta$$

$$\sigma^2(\hat{\vartheta}_{MV}) = \text{Var}(\hat{\vartheta}_{MV}) \to \left[-\frac{\partial^2 L(\hat{\vartheta}_{MV})}{\partial \vartheta^2} \right]^{-1}$$

por tanto, el error relativo de estimación de ϑ, dado por:

$$\frac{\vartheta - \hat{\vartheta}_{MV}}{\sigma(\hat{\vartheta}_{MV})}$$

sigue una distribución normal estándar, y podemos construir el intervalo:

$$\hat{\vartheta}_{MV} - z_{\alpha/2}\sigma(\hat{\vartheta}_{MV}) \leq \vartheta \leq \hat{\vartheta}_{MV} + z_{\alpha/2}\sigma(\hat{\vartheta}_{MV}) \tag{8.22}$$

Un inconveniente de este método general es que la convergencia de la distribución de $\hat{\vartheta}_{MV}$ hacia la normal puede ser muy lenta y entonces el intervalo (8.22) será poco preciso. Esto no ocurre cuando ϑ es un parámetro de centralización, y los intervalos (8.8), (8.10), (8.11), (8.12), (8.17) y (8.20) son casos particulares donde este método funciona satisfactoriamente. Para varianzas, la distribución de log s^2 suele ser más simétrica que la de s^2, y se recomienda construir el intervalo suponiendo normalidad sobre el logaritmo.

Ejemplo 8.4

Se han observado cuatro elementos defectuosos de un total de 200 examinados entre los producidos por un proceso. Construir su intervalo de confianza aproximado para la proporción de elementos defectuosos en la fabricación utilizando las propiedades asintóticas del estimador *MV*.

El estimador máximo-verosímil de p, proporción de elementos defectuosos en la población, es $\hat{p}$, proporción observada. Por otro lado, por ser estimador máximo-verosímil, $\hat{p}$ será asintóticamente normal, con media p y varianza:

$$\text{Var}(\hat{p}) = \left[-\frac{\partial^2 L(\hat{p})}{\partial p^2} \right]^{-1}$$

Para escribir la función soporte, sean x_i ($i = 1, \ldots, 200$) las variables de la muestra, donde supondremos que $x_i = 1$, si el elemento es defectuoso, y $x_i = 0$ en otro caso. Entonces (véase el ejemplo 7.11)

$$L(p) = \Sigma x_i \ln p + (n - \Sigma x_i) \ln (1 - p)$$

y la tasa de discriminación es:

$$\frac{\partial L(p)}{\partial p} = \frac{\Sigma x_i}{p} - \frac{n - \Sigma x_i}{1 - p}$$

que, igualada a cero, proporciona el estimador *MV*: $\hat{p} = \Sigma x_i/n$. Derivando de nuevo:

$$\frac{\partial^2 L(p)}{\partial p^2} = \frac{-\Sigma x_i}{p^2} - \frac{n - \Sigma x_i}{(1 - p)^2}$$

y, sustituyendo $\hat{p} = \Sigma x_i/n = r/n$.

$$\left[-\frac{\partial^2 L(\hat{p})}{\partial p^2} \right]^{-1} = \left[\frac{n^2}{r} + \frac{n^2}{n - r} \right]^{-1} = \frac{r(n - r)}{n^3} = \frac{\hat{p}\hat{q}}{n}$$

Por tanto, el estadístico

$$\frac{p - \hat{p}}{\sqrt{\frac{\hat{p}\hat{q}}{n}}}$$

tendrá, asintóticamente, una distribución $N(0, 1)$. Un intervalo aproximado será:

$$p \in \left(\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}} \right)$$

que coincide con el introducido como caso particular de la media en (8.12). Tomando $\alpha = 0, 1$, como $z_{\alpha/2} = 1{,}64$ el intervalo resultante será:

$$p \in \left(\frac{4}{200} \pm 1{,}64 \sqrt{\frac{4 \cdot 196}{200^3}} \right)$$

es decir:

$$p \in (0{,}02 \pm 0{,}016)$$

y por tanto:

$$0{,}004 < p < 0{,}036$$

y la proporción defectuosa está entre el 4 por mil y el 3,6 por cien.

8.10 Determinación del tamaño muestral

Las fórmulas deducidas para los intervalos de confianza nos permiten deducir el tamaño muestral necesario para obtener una precisión determinada. Veamos algunos ejemplos:

Media

Si se desea que el intervalo de confianza $1 - \alpha$ tenga una amplitud $\bar{x} \pm L$, tendremos que

$$L = z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \tag{8.23}$$

lo que exige un valor de *n:*

$$\boxed{n = \frac{z_{\alpha/2}^2 \sigma^2}{L^2}} \tag{8.24}$$

Observemos que esto exige el conocimiento de σ. Cuando σ es desconocido, tendremos que tomar una *muestra piloto* pequeña y estimar σ mediante $\hat{s}$. Por ejemplo, si $L = 2$, $\hat{s} = 7$ y $\alpha = 0{,}05$ ($z_{\alpha/2} = 1{,}96$) el tamaño muestral necesario es 47.

Proporciones

Si deseamos que el intervalo sea del tipo $\hat{p} \pm L$, aplicando el razonamiento anterior:

$$L = z_{\alpha/2}\sqrt{\frac{pq}{n}}$$

como p es desconocido, podemos ponernos en la situación más desfavorable, $p = 0{,}5$, y obtener:

$$\boxed{n = \frac{z^2_{\alpha/2}}{4\,L^2}} \qquad (8.25)$$

Ejemplo 8.5

Calcular qué tamaño de muestra se debe tomar para estimar las diferencias entre los votos de dos partidos *A* y *B*, si se desea que el intervalo del 95% sea del tipo $d \pm 0{,}02$, donde $d = \hat{p}_A - \hat{p}_B$, suponiendo que los votantes pueden elegir entre muchos partidos políticos.

Aunque estrictamente es claro que las estimaciones de los votos de los partidos no son independientes, ya que la suma de todos los votos debe ser el 100%, si suponemos que hay muchos partidos podemos tomar $\hat{p}_A$ y $\hat{p}_B$ como aproximadamente independientes. Entonces:

$$p_A - p_B \in \left(\hat{p}_A - \hat{p}_B \pm 1{,}96\sqrt{\frac{\hat{p}_A\hat{q}_A}{n} + \frac{\hat{p}_B\hat{q}_B}{n}}\right)$$

Obtenemos un tamaño muestral máximo colocándonos en el caso más desfavorable, $\hat{p}_A = \hat{p}_B = 0{,}5$ tendremos:

$$0{,}02 = 1{,}96\sqrt{\frac{2 \cdot 0{,}5^2}{n}}$$

$$n = \frac{1{,}96^2 \cdot 2 \cdot 0{,}25}{0{,}02^2} = 4802$$

El tamaño de casi 5.000 personas para dilucidar el resultado (si suponemos que 2 puntos de votos en diferencia garantizan la mayoría de escaños en el parlamento) es alto, porque nos hemos puesto en el caso extremo. Si, por resultados anteriores, podemos suponer que el voto de estos partidos será, como máximo, del 35%, entonces tomando $\hat{p}_A = \hat{p}_B = 0{,}35$, tenemos que

$$n = \frac{1{,}96^2 \cdot 2 \cdot 0{,}35^2}{0{,}02^2} = 2353$$

que es el tamaño de muestra utilizado en muchos sondeos electorales.

8.11 La estimación autosuficiente de intervalos de confianza *(bootstrap)*

8.11.1 Introducción

La media muestral es un estimador con una propiedad muy especial: su precisión, medida por la inversa de la varianza de su distribución muestral, puede conocerse fácilmente a partir de la varianza de la muestra:

$$\hat{S}^2(\bar{x}) = \frac{\hat{S}^2}{n} = \left[\frac{1}{n(n-1)}\Sigma(x_i - \bar{x})^2\right]$$

y además, esta expresión es válida en general, no dependiendo del modelo de distribución de probabilidad que genera la muestra. Al estimar otra característica cualquiera de la población, como la varianza, el coeficiente de asimetría, el de curtosis o cualquier otro parámetro, la precisión de la estimación depende de la distribución que genera los datos. Si utilizamos un estimador de máxima verosimilitud podemos conocer su varianza asintótica, pero esta medida puede ser muy poco precisa en muestras pequeñas y depende, además, de la hipótesis sobre la distribución.

Los métodos de *estimación herramental (jackknife)* y *estimación autosuficiente (bootstrap)* son métodos generales para obtener la precisión de un estimador de forma aproximada sin hacer hipótesis respecto a su distribución. Ambos proporcionan respuestas rápidas a problemas de difícil tratamiento algebraico, requiriendo en contrapartida el uso extensivo del ordenador. Su uso actual ha sido posible por la potencia y rapidez de los ordenadores digitales.

El método herramental fue desarrollado por Quenouille en 1949 como procedimiento para reducir el sesgo de un estimador y bautizado por Tukey en 1958 que lo denominó *jackknife* (literalmente, navaja de usos múltiples), generalizándolo como método general de estimación. La estimación autosuficiente es debida a Efron (véase Efron, 1982), que la denominó *bootstrap* (literalmente, correas que ayudan a calzarse las botas) haciendo referencia a una expresión anglosajona (levantarse tirando de las propias correas de las botas) que refleja la autosuficiencia del método.

Vamos a exponer en esta sección el método autosuficiente *(bootstrap)*, que tiene la ventaja de la generalidad y sencillez de cálculo. En el apéndice 8A presentamos el método herramental.

8.11.2 La estimación autosuficiente *(bootstrap)*

Este método se basa en calcular directamente la varianza del estimador considerando la muestra como si fuese toda la población y aplicando el método de Montecarlo para obtener réplicas de la muestra. En concreto, dada una muestra $(x_1, ..., x_n)$, el método procede como sigue:

1) Considerar la muestra como una población de una variable que toma los n valores posibles $(x_1, ..., x_n)$ con probabilidad $1/n$. Extraer una muestra aleatoria simple de tamaño n de dicha población mediante el método de Montecarlo. Esto equivale a obtener una muestra al azar con reemplazamiento de los valores observados. Esta muestra generada no coincidirá, en general, con la muestra original. Sea $(y_1^{[1]}, ..., y_n^{[1]})$ la muestra así obtenida.
2) Calcular en la muestra generada en el paso anterior el estimador $\hat{\vartheta}_1 = \hat{\vartheta}(y_1^{[1]}, ..., y_n^{[1]})$ cuya precisión queremos estimar.
3) Repetir los pasos 1) y 2) un número B grande de veces (1.000 por ejemplo). Obtendremos así una secuencia de B valores del estimador, $\hat{\vartheta}_1, ..., \hat{\vartheta}_B$ que consideraremos la distribución de valores de $\hat{\vartheta}$. Su media será:

$$\hat{\vartheta}_m = \frac{1}{B}\Sigma\hat{\vartheta}_i$$

y su varianza:

$$\text{Var}(\hat{\vartheta}) = \frac{1}{B}\Sigma(\hat{\vartheta}_i - \hat{\vartheta}_m)^2$$

Puede demostrarse que, en condiciones generales, este método obtiene asintóticamente la varianza del estimador $\hat{\vartheta}$, y que el intervalo de confianza de nivel $1 - \alpha$ puede obtenerse de la distribución de los B valores de $\hat{\vartheta}_i$. Para ello se obtienen dos valores $\hat{\vartheta}_{INF}$ y $\hat{\vartheta}_{SUP}$ tales que:

$$P(\hat{\vartheta}_{INF} \leq \hat{\vartheta}_i \leq \hat{\vartheta}_{SUP}) = 1 - \alpha$$

Entonces $(\hat{\vartheta}_{INF}, \hat{\vartheta}_{SUP})$ proporciona un intervalo de confianza de nivel $1 - \alpha$. Estos límites se calculan ordenando los valores $\hat{\vartheta}_i$ y tomando $\hat{\vartheta}_{INF}$ y $\hat{\vartheta}_{SUP}$ como los valores situados en las posiciones $[B \times \alpha/2]$ y $[B \times (1 - \alpha/2)]$, donde [ ] indica redondear el entero más próximo.

Como ilustración vamos a comprobar que este método proporciona la respuesta correcta en la estimación de la media. Llamando ahora $\hat{\vartheta}_i = \bar{y}$, el valor esperado de este estadístico es:

$$E[\hat{\vartheta}_i] = \sum_{i=1}^{n} x_i p(x_i) = \sum_{i=1}^{n} x_i/n = \bar{x}$$

y comprobamos que el valor medio coincide con la media muestral. Su varianza teórica será σ^2/n, siendo σ^2 la varianza de la distribución base, que es:

$$\sigma^2 = \text{Var}(y) = \Sigma\,(x_i - \bar{x})^2 p(x_i) = \frac{1}{n}\,\Sigma(x_i - \bar{x})^2$$

Por tanto:

$$\text{Var}(\hat{\vartheta}) = \frac{1}{n^2}\,\Sigma\,(x_i - \bar{x})^2$$

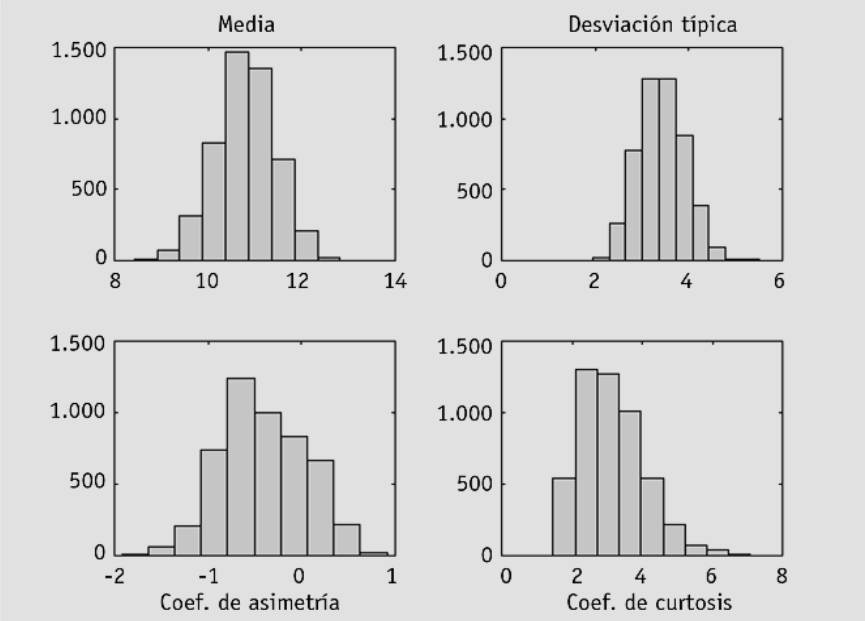
que es, asintóticamente, el resultado que se obtendría por la estimación autosuficiente. En el ejercicio 8.5 se comprueba esta propiedad experimentalmente, y en el apéndice 8B se explica cómo utilizar un programa que proporcione números aleatorios para construir los intervalos de confianza autosuficientes.

Ejemplo 8.6

Vamos a comprobar que la estimación autosuficiente proporciona valores próximos a los exactos para estimar los parámetros de una población normal. Para ello vamos a generar los datos muestrales de una distribución $N(10,3)$ (que naturalmente en la práctica sería desconocida), para comprobar cómo funciona el método autosuficiente. Tomando por Montecarlo 30 valores al azar de esta distribución, la muestra observada sería: 13,9443, 11,9960, 9,1747, 9,9309, 7,2761, 6,8690, 11,1205, 12,7046, 13,8356, 9,6146, 11,8385, 15,8696, 16,7990, 8,8781, 16,7141, 9,5213, 7,8902, 11,6904, 9,8491, 13,4908, 11,9764, 5,3497, 0,9126, 11,6217, 6,9730, 12,7241, 14,7469, 7,0627, 13,0237 y 10,4755. La media de estos 30 datos es 10,7958 y la desviación típica corregida por grados de libertad es 3,5227.

Aplicamos ahora el método autosuficiente a esta muestra y tomamos 5.000 muestras con reemplazamiento, cada una de tamaño 30, de la población de 30 valores que forma la muestra original. En cada una de estas muestras calculamos la media, la desviación típica, el coeficiente de asimetría y el de apuntamiento. La figura 8.5 proporciona la distribución de los valores obtenidos para estos estadísticos.

Figura 8.5 Generación de distribuciones autosuficientes para la media, la desviación típica y los coeficientes de asimetría y curtosis



Se observa que, en este caso, las distribuciones autosuficientes de la media y la desviación típica son aproximadamente simétricas, mientras que las de los coeficientes de asimetría y curtosis no lo son. Cada una de estas distribuciones proporciona automáticamente un intervalo de confianza para el parámetro correspondiente. La tabla 8.1 presenta los intervalos construidos bajo la hipótesis de normalidad para estos datos y los obtenidos con la estimación autosuficiente. Puede comprobarse que la aproximación es muy buena.

Tabla 8.1 Resultados de los intervalos de confianza bajo normalidad y con el método autosuficiente

	Normalidad	Autosuficiente
Media	9,48; 12,11	9,52; 12,04
Desv. típica	2,80; 4,74	2,49; 4,44
C. asimetr.	–1,40; 0,35	–1,21; 0,43
C. curtosis	1,73; 5,24	1,85; 5,11

Los intervalos de la tabla 8.1 se han calculado de la forma siguiente. En la columna de normalidad el intervalo para la media es

$$\bar{x} \pm t_{29}(\alpha/2)\frac{\hat{s}}{\sqrt{n}} = 10{,}7958 \pm 2{,}04\,\frac{3{,}5227}{\sqrt{30}}$$

el intervalo para la desviación típica se ha calculado tomando la raíz cuadrada del intervalo para la varianza:

$$\left(\sqrt{\frac{29\hat{s}^2}{\chi^2_a}}, \sqrt{\frac{29\hat{s}^2}{\chi^2_b}}\right) = \left(\sqrt{\frac{29[3{,}5227]^2}{45{,}7}}, \sqrt{\frac{29[3{,}5227]^2}{16}}\right)$$

Finalmente, para los coeficientes de asimetría y curtosis se ha utilizado que son asintóticamente normales con desviaciones típicas $\sqrt{6/n}$ y $\sqrt{24/n}$ respectivamente.

En la columna autosuficiente de la tabla 8.1 los intervalos autosuficientes del 95% se calculan, en todos los casos, ordenando los 5.000 valores calculados para cada estadístico en cada una de las 5.000 muestras con reemplazamiento generadas y tomando como extremos del intervalo los valores situados en las posiciones 5000*.025 = 125 y 5000*.975 = 4875.

Ejemplo 8.7

Los siguientes 25 datos corresponden a una prueba de capacidad utilizada por una empresa para seleccionar aspirantes: 47, 37, 71, 69, 70, 62, 65, 57, 60, 68, 77, 45, 82, 54, 39, 59, 47, 75, 52, 36, 71, 69, 24, 73 y 66. La media de estos datos es 59,04 y la desviación típica corregida es 14,82. El coeficiente de asimetría es –0,61 y el de curtosis 2,50. El histograma de estos datos se presenta en la figura 8.6 y existen dudas de que la distribución sea normal. Vamos a utilizar el método autosuficiente para calcular un intervalo de confianza para la media, la desviación típica, el coeficiente de asimetría y de curtosis de la población que genera estos datos.

La figura 8.7 presenta las distribuciones obtenidas con el método autosuficiente. De estas distribuciones obtenemos los intervalos: media (52,97; 64,37), desviación típica (10,64, 17,89), coeficiente de asimetría (–1,27; 0,05), coeficiente de apuntamiento o curtosis (1,60; 4.27).

Figura 8.6 Histograma de los datos del ejemplo 8.6

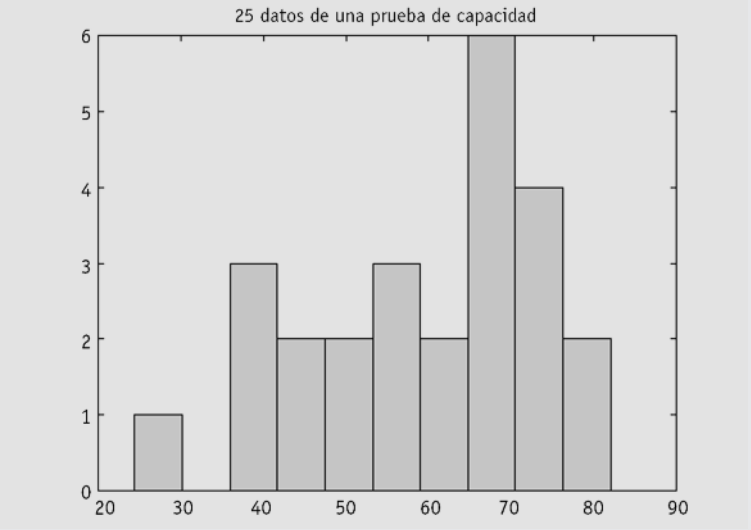
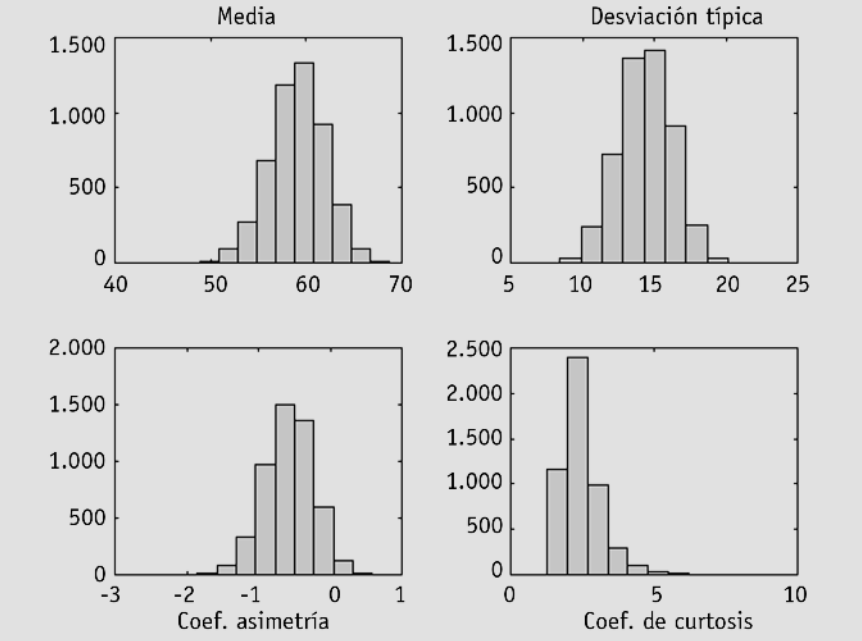


Figura 8.7 Distribuciones obtenidas por el método autosuficiente



Ejercicios 8

8.1. A continuación se indica la edad en que diez importantes matemáticos hicieron su primer descubrimiento fundamental. Utilizar esta muestra para estimar la edad a la que un matemático producirá su primera contribución fundamental. Descartes (23), Fermat (27), Pascal (31), Newton (23), Leibniz (29), Lagrange (23), Laplace (24), Galois (21), Gauss (18), Poincaré (28).

8.2. En la lista adjunta se indica la edad y el área científica en que trece importantes científicos de diversas áreas descubrieron la teoría que les ha dado la fama. Construir con estos datos un intervalo de confianza para la edad a la que los científicos realizan su contribución más importante: Galileo (34, astronomía), Franklin (40, electricidad), Lavoiser (31, química), Lyell (33, geología), Darwin (49, biología) Maxwell (33, ecuaciones de la luz), Curie (34, radioactividad), Plank (43, teoría cuántica), Marx (30, socialismo científico), Freud (31, psicoanálisis), Bohr (26, modelo del átomo), Einstein (26, relatividad), Keynes (36, macroeconomía).

8.3. Construir el intervalo de confianza para la diferencia entre la edad promedio a la que los matemáticos y los científicos en general hacen su contribución fundamental, suponiendo que la variabilidad es la misma en ambas poblaciones.

8.4. Construir el intervalo anterior pero sin suponer que las varianzas son iguales.

8.5. Una muestra de 40 canciones emitidas por una cadena de radio durante una semana conduce a que la duración media por canción es de 3,4 minutos con una desviación típica de 1,2 minutos. Calcular un intervalo de confianza para la duración media de las canciones emitidas por dicha emisora.

8.6. Una muestra de 12 estaciones de servicio de una cadena de gasolineras proporciona un ingreso medio por persona al mes de 2340 euros con una desviación típica de 815 euros. Calcular un intervalo de confianza para el ingreso medio por trabajador en esta empresa.

8.7. Calcular el número de estaciones que debemos estudiar en el problema anterior para que el intervalo tenga un amplitud máxima de 500 euros.

8.8. Un banco realiza una encuesta para determinar la proporción de clientes satisfechos con un servicio. En la sucursal A con una muestra de 100 personas se han obtenido 76 satisfechos mientras que en la B una muestra de 140 personas obtiene 112 personas satisfechas. Construir un intervalo de confianza para las diferencias entre las satisfacciones medias entre ambas sucursales.

8.9. Para estimar la media de una $N(\mu, \sigma^2)$ con σ^2 conocido y una muestra de tamaño n, determinar el tamaño de n para que el intervalo del 0,99 para μ tenga longitud L.

8.10. Una muestra de tamaño 10 de una $N(\mu_1, 225)$ resulta con $\bar{x}_1 = 170{,}2$ y otra de tamaño 12 de $N(\mu_2, 256)$ conduce a $\bar{x}_2 = 176{,}7$. Calcular un intervalo del 95% para $\mu_1 - \mu_2$.

8.11. Dos muestras de dos poblaciones normales han dado los siguientes resultados: $n_1 = 8$, $\Sigma x_i = 12$; $\Sigma x_i^2 = 46$; $n_2 = 11$; $\Sigma y_i = 22$; $\Sigma y_i^2 = 80$. Calcular un intervalo de confianza para σ_1^2/σ_2^2 al 95%.

8.12. La tensión entre bornes de las baterías de cierta marca a la salida de fábrica es $4 \pm e$ voltios, donde e tiene una distribución uniforme entre (–0,25; +0,25) voltios. Se conectan 25 baterías en serie. Calcular un intervalo de confianza para la fuerza electromotriz total con $\alpha = 0{,}05$.

8.13. Una encuesta de l00 votantes para conocer las opiniones respecto a dos candidatos muestra que 55 apoyan a A y 45 a B. Se pide:

a) Calcular un intervalo de confianza para la proporción de votos de cada candidato.
b) Calcular cuál debería haber sido el tamaño muestral para que una fracción 0,55 de partidarios de A permita afirmar que será elegido al 95%.

8.14. Se realizan diez determinaciones del porcentaje de riqueza en un polímero con dos instrumentos distintos. Las varianzas muestrales resultan ser 0,5919 y 0,6065. Encontrar un intervalo de confianza para el cociente de varianzas teóricas en ambos instrumentos.

8.15. Se estudian dos tipos de neumáticos con los resultados siguientes: tipo A: $n_1 = 121$; $\bar{x}_1 = 27.465$ km; $\hat{s}_1 = 2.500$ km. Tipo B: $n_2 = 121$; $\bar{x}_2 = 27.572$ km; $\hat{s}_2 = 3.000$ km. Calcular, con $\alpha = 0{,}01$:

a) Un intervalo de confianza para σ_1^2/σ_2^2.
b) Un intervalo de confianza para $\mu_1 - \mu_2$.

8.16. Una compañía contrata 10 tubos con filamentos de tipo A y diez con filamentos de tipo B. Las duraciones de vida observadas han sido:

A: 1614; 1094; 1293; 1643; 1466; 1270; 1340; 1380; 1028; 1497.
B: 1383; 1138; 1092; 1143; 1017; 1061; 1627; 1021; 1711, 1065.

a) Suponiendo que las varianzas son iguales, encontrar un intervalo de confianza para la diferencia de medias.
b) Lo mismo pero suponiendo las varianzas desiguales.

8.17. Obtener un intervalo asintótico para:

a) El parámetro λ en una distribución de Poisson.
b) El parámetro λ en una distribución exponencial.

8.18. Se comparan las producciones de dos máquinas A y B, que fabrican elementos en serie. En una muestra de 200 elementos de A, resultaron 16 defectuosas, mientras que en otra de 100 de B resultaron 12 defectuosas. Calcular:
   a) Un intervalo para la diferencia de proporción defectuosa en ambas máquinas con $\alpha = 0{,}05$.
   b) Dibujar la distribución de confianza.

8.19. En un estudio sobre la afectividad de los estudiantes universitarios se pregunta a 20 personas sobre el número de personas del sexo opuesto con el que ha mantenido relaciones afectivas durante los tres últimos años. Los resultados obtenidos han sido (1, 0, 2, 0, 3, 3, 1, 5, 1, 2, 0, 0, 1, 0, 4, 2, 1, 0, 6, 2, 1, 1, 2, 1, 1). Obtener intervalos de confianza del 95% para la media y la desviación típica utilizando el método autosuficiente (véase el apéndice 8B).

8.20. Comparar el intervalo de confianza autosuficiente con el asintótico para la media de una distribución exponencial.

8.12 Resumen del capítulo y consejos de cálculo

En este capítulo hemos visto primero cómo construir intervalos de confianza conociendo la distribución que genera los datos y después sin conocerla. La base del primer método es encontrar un estadístico pivote que tenga una distribución conocida, salvo por el parámetro que se desea estimar. La base del segundo método es la generación de muestras mediante el método autosuficiente. El primer método es más rápido y simple, pero no puede aplicarse siempre. El segundo es totalmente general y nos proporciona una respuesta automática en muchos problemas complejos.

El cuadro 8.1 resume la construcción de los intervalos de confianza más comunes. Para problemas más complejos puede acudirse a los resultados asintóticos de los estimadores *MV* o a los intervalos autosuficientes.

El resultado más importante de este capítulo es que, para muestras grandes, en condiciones muy generales:

$$\frac{\text{estimador-parámetro}}{\text{desviación típica del estimador}} = N(0,1)$$

Para medias de muestras pequeñas de poblaciones normales:

$$\frac{\text{media estimada - media verdadera}}{\text{desviación típica estimada}} = t_g$$

Cuadro 8.1 Resumen de intervalos de confianza

a) *Proporciones*

Parámetro	Estadístico pivote	Distribución	Intervalo de confianza
p	$(\hat{p}-p)/\sqrt{\hat{p}\hat{q}/n}$	$N(0,1)$	$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$
$p_1 - p_2$	$\frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}}$	$N(0,1)$	$\hat{p}_1-\hat{p}_2 \pm \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}$

b) *Poblaciones normales*

Parámetro	Estadístico pivote	Distribución	Intervalo de confianza
μ, σ conocido	$\sqrt{n}(\bar{x}-\mu)/\sigma$	$N(0,1)$	$\bar{x} \pm z_{\alpha/2}\sigma/\sqrt{n}$
μ, σ desconocido	$\sqrt{n}(\bar{x}-\mu)/\hat{s}$	t_{n-1}	$\bar{x} \pm t_{\alpha/2}\hat{s}/\sqrt{n}$
σ^2	$(n-1)\hat{s}^2/\sigma^2$	χ^2_{n-1}	$(n-1)\hat{s}^2/\chi^2_{\alpha/2}$, $(n-1)\hat{s}^2/\chi^2_{1-\alpha/2}$
$\mu_1-\mu_2$ $\sigma_1=\sigma_2$	$\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\hat{s}_T\sqrt{n_1^{-1}+n_2^{-1}}}$	$t_{n_1+n_2-2}$	$\bar{x}_1-\bar{x}_2 \pm t_{\alpha/2}\hat{s}_T\sqrt{n_1^{-1}+n_2^{-1}}$
$\mu_1-\mu_2$ $\sigma_1\neq\sigma_2$	$\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\hat{s}_1^2/n_1+\hat{s}_2^2/n_2}}$	$t_{n_1+n_2-\Delta-2}$	$\bar{x}_1-\bar{x}_2 \pm t_{\alpha/2}\sqrt{\hat{s}_1^2/n_1+\hat{s}_2^2/n_2}$

c) *Intervalos asintóticos*

Parámetro	Estadístico pivote	Distribución	Intervalo de confianza
ϑ	$\frac{\vartheta-\hat{\vartheta}_{MV}}{\sigma(\hat{\vartheta}_{MV})}$	$N(0,1)$	$\hat{\vartheta}_{MV} \pm z_{\alpha/2}\sigma(\hat{\vartheta}_{MV})$

y para la varianza de poblaciones normales:

$$\frac{\text{suma de desviaciones al cuadrado}}{\text{varianza de la población}} = x_g^2$$

donde g es el número de residuos independientes (grados de libertad).

Los programas estadísticos proporcionan directamente mediante un comando los intervalos de confianza que hemos estudiado. Sin embargo, no es habitual que proporcionen los intervalos autosuficientes, aunque éstos pueden calcularse fácilmente con un pequeño trabajo adicional. En el apéndice 8B hemos detallado cómo construir los intervalos de confianza autosuficientes con varios programas de uso habitual.

8.13 Lecturas recomendadas

La estimación por intervalos se trata en todos los manuales de estadística básica que se listan en la bibliografía. Guttman *et al.* (1982) y Hogg y Ledolter (1992) contienen aplicaciones a la ingeniería, y Wonnacott y Wonnacott (2004), Newbold *et al.* (2006) y Webster (2005) a la economía.

Una excelente exposición del método autosuficiente se encuentra en Efron y Tibshirani (1994). Véase también Efron (1987).

Apéndice 8A: El método herramental *(jackknife)*

Supongamos que estimamos ϑ mediante un estimador definido $\hat{\vartheta}$. Vamos a presentar el método herramental para obtener una estimación de su varianza muestral.

1) Eliminar de la muestra uno cualquiera de los n valores muestrales, x_i, y calcular el valor del estimador en la muestra de tamaño $n - 1$. Llamaremos $\mathbf{X}_{(i)}$ a la muestra sin el elemento x_i y $\hat{\vartheta}_{(i)}$ al estimador obtenido con dicha muestra. Repitamos este procedimiento n veces para obtener estimadores $\hat{\vartheta}_{(1)}$, $\hat{\vartheta}_{(2)}$, $\hat{\vartheta}_{(3)}$, ..., $\hat{\vartheta}_{(n)}$, obtenidos eliminando el primero, segundo, ..., n-ésimo término y calculando el estimador en la muestra restante de $n - 1$ elementos. Por ejemplo, para el estimador $\bar{x}$ de μ:

$$\hat{\vartheta}_{(i)} = \bar{x}_{(i)} = \frac{1}{n-1} \sum_{j \neq i} x_j$$

y para el estimador $\hat{s}^2$ de σ_2:

$$\hat{\vartheta}_{(i)} = \hat{s}^2_{(i)} = \frac{1}{n-2} \sum_{j \neq i} (x_j - \bar{x}_{[j]})^2$$

2) Puede demostrarse que la varianza de $\hat{\vartheta}$ se aproxima por:

$$s^2(\hat{\vartheta}) = \frac{n-1}{n} \Sigma(\hat{\vartheta}_{[i]} - \hat{\vartheta}_{[\cdot]})^2$$

donde:

$$\hat{\vartheta}(\cdot) = \frac{1}{n} \Sigma \hat{\vartheta}_{(n)}$$

Por ejemplo, la varianza de la media muestral $\bar{x}$ se aproxima por:

$$s^2(\bar{x}) = \frac{n-1}{n} \Sigma(\bar{x}_{[i]} - \bar{x})^2$$

y como:

$$\bar{x}_{(i)} = \frac{n}{n-1} \bar{x} - \frac{1}{n-1} x_i$$

se obtiene:

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

que es la expresión exacta.

Para construir intervalos de confianza utilizaremos un estimador modificado que tiene siempre menor sesgo que $\hat{\vartheta}$ y la misma varianza. Definamos los pseudovalores muestrales por:

$$d_{(i)} = \hat{\vartheta} + (n-1)\,(\hat{\vartheta} - \hat{\vartheta}_{[i]})$$

Entonces, el estimador

$$d(\cdot) = \frac{1}{n} \Sigma d_{(i)} = \hat{\vartheta} + (n-1)\,(\hat{\vartheta} - \hat{\vartheta}_{[\cdot]})$$

donde $\hat{\vartheta}_{(\cdot)} = \Sigma\hat{\vartheta}_{(i)}/n$ como antes, es un estimador con menor sesgo que $\hat{\vartheta}$. Su varianza se calcula como sigue: la varianza muestral de $d_{(1)}, ..., d_{(n)}$ es:

$$s^2(d) = \frac{1}{n-1}\Sigma(d_{[i]} - d_{[\cdot]})^2 = (n-1)\,\Sigma\,(\hat{\vartheta}_{[i]} - \hat{\vartheta}_{[\cdot]})^2$$

Por tanto, la varianza muestral de $d_{(\cdot)}$ será:

$$\text{Var}\,(d_{[\cdot]}) = \frac{s^2(d)}{n} = \frac{n-1}{n}\Sigma(\hat{\vartheta}_{[i]} - \hat{\vartheta}_{[\cdot]})^2$$

que coincide con la antes obtenida. Podemos construir intervalos de confianza utilizando el teorema central del límite: asintóticamente

$$\frac{d_{(\cdot)} - \vartheta}{\sqrt{\text{Var}\,(d_{[\cdot]})}} \text{ es } N(0, 1).$$

Por tanto, un intervalo de confianza para ϑ será:

$$\vartheta \in \bar{d} \pm 1{,}96\sqrt{\text{Var}\,(d)}$$

Apéndice 8B: Construcción mediante ordenador de intervalos de confianza por el método autosuficiente

Cualquier programa informático que contenga la generación de números aleatorios y la posibilidad de realizar bucles y cálculos simples puede utilizarse para calcular estos intervalos. En particular, Matlab, Gauss, S-plus, Sca, Minitab y Sas, entre otros programas, tienen esta capacidad. También puede realizarse con programas como Statgraphics o Excel, pero entonces en lugar de programar un bucle hay que seleccionar todas las muestras de golpe y después calcular los estadísticos sobre grupos consecutivos de n observaciones.

Vamos a ilustrar los principios generales con un ejemplo con Matlab. Generaremos 500 muestras (tomamos B = 500) con reemplazamiento de un vector de datos de tamaño 30, (n = 30), calcularemos la media y la desviación típica de cada muestra y el intervalo de confianza autosuficiente para cada parámetro. Suponemos que la muestra inicial está en un vector columna x de dimensiones 30×1.

1. Comenzamos definiendo un bucle para generar 500 veces una muestra y calcular la media y desviación típica. En Matlab el bucle se define mediante la instrucción

for i = 1:500

2. A continuación calculamos un vector de 30 valores que va a indicar cuáles de los componentes del vector de datos *x* van a tomarse en cada muestra generada. Para ello se generan 30 números aleatorios con la instrucción

   rand(30,1)

   que proporciona un vector de 30 × l valores uniformes entre 0 y 1. Como necesitamos números enteros entre 1 y 30, primero los multiplicamos por 30, para obtener números reales entre cero y 30 [la instrucción es 30*rand (30,1)] y después los redondeamos hacia arriba para obtener números aleatorios enteros uniformes entre 1 y 30. Esto se hace en Matlab con la instrucción ceil, con lo que la instrucción completa que nos proporciona el vector de datos es

   id = ceil[30*rand(30,1)]

   Observemos que la instrucción round (redondear al entero más próximo) no sería adecuada, porque obtendríamos valores entre los enteros 0,30 pero donde los valores 0 y 30 tienen la mitad de probabilidad del resto de números. Si utilizásemos esta instrucción tendríamos que transformar los 0 obtenidos por 30, para tener números equiprobables entre 1 y 30.
3. Elegiremos la muestra generada tomando las observaciones del vector **x** definidas por los indicadores id generados en el paso anterior y guardaremos la muestra de tamaño 30 así generada en un vector **xn**. La instrucción es:

   xn = x(id)

4. Calcularemos ahora la media de la muestra generada en el paso anterior con la instrucción mean y guardaremos el resultado en un vector **m**. El indicador *i* variará de 1 hasta 500 al estar dentro del bucle:

   m(i) = mean(xn)

5. Calcularemos la desviación típica con la instrucción de Matlab std y guardaremos el resultado en des:

   des(i) = std(xn)

6. Finalizamos el bucle. De esta manera se realizarán 500 repeticiones de las instrucciones 2 a 5:

   end

   El resultado de este bucle son 500 valores de la media y la desviación típica muestrales que pueden dibujarse en un histograma y utilizarse para calcular intervalos de confianza para la media y desviación de la población. Para ello, se utilizan las instrucciones siguientes en Matlab.
7. Ordenamos los valores de las medias, m, y los colocamos en otro vector sm:

   sm = sort(m)

8. Lo mismo para las desviaciones típicas

   sdes = sort(des)

9. Se calcula el índice del valor ordenado inferior multiplicando 1 – $\alpha/2$ por B y redondeando al entero más próximo el resultado:

   h1 = round([(alfa/2)*B])

10. Lo mismo para el valor ordenado inferior multiplicando 1–$\alpha/2$ por B y redondeando el resultado al entero más próximo:

    h2 = round([(l–(alfa/2)]*B))

11. Se indica el intervalo para la media:

    [sm(h1) sm(h2)]

12. Y para la desviación típica:

    [sdes(hl) sdes(h2)]

El análisis con Gauss es prácticamente el mismo. Efron y Tibshirani (1993) incluyen los programas para S-plus. Algunos programas, como Minitab, permiten directamente obtener una muestra con reemplazamiento de los datos. Por ejemplo, si los datos están en un vector c1 podemos obtener una muestra con reemplazamiento de tamaño *n* y almacenarla en c2 con la instrucción:

MTB > Sample *n* cl c2

SUBC> Replace

Estas instrucciones pueden integrarse dentro de un macro con un DO para repetirlas el número de muestras deseadas. Statgraphics proporciona directamente *n* valores uniformes enteros entre *a* y *b* con la instrucción: rinteger(n,a,b). Sin embargo, como con este programa no se pueden programar bucles, habrá que generar en una columna los B × *n* números aleatorios que vamos a necesitar [rinteger(B × *n*, *a*, *b*)], sustituir cada valor aleatorio de esta columna por el correspondiente de la muestra (el número 3 de esta columna indica el tercer elemento muestral, etc.) con la instrucción RECODING DATA y a continuación calcular el estadístico de interés con grupos de tamaño *n* de esta columna con la instrucción SELECT. Por supuesto ésta es una de las formas posibles, pero el lector experto en este programa puede explorar otras formas alternativas. Excel proporciona también números aleatorios, por lo que puede utilizarse de una forma similar a Statgraphics. Si la muestra está en las posiciones 1 a *n* de la columna A la siguiente instrucción de Excel genera muestras, una muestra aleatoria con reemplazamiento de los datos de la columna A en las posiciones A1 hasta An:

```
=INDICE($A$1:$A$n;REDONDEAR.MAS(ALEATORIO()*n;0);1)
```

La instrucción INDICE requiere tres argumentos. El primero es una matriz de datos de donde vamos a tomar valores. El segundo es el número de fila que vamos a seleccionar de esa matriz de datos. El tercero es el número de columna que especifiquemos. Estos tres argumentos están separados por ;. La matriz es, en este caso, la columna donde están los datos entre las posiciones 1 a n ($A$1:An). Para seleccionar una fila al azar, primero generamos un número aleatorio uniforme entre cero y uno con la instrucción ALEATORIO(), después lo multiplicamos por *n* para tenerlo entre cero y n, a continuación redondeamos hacia arriba para que los valores estén entre 1 y *n* y éste es el indicador de la fila a seleccionar. La columna siempre es la misma, la columna 1. Copiando esta instrucción podemos generar tantas muestras como queramos. Una vez generadas las muestras aplicamos la función correspondiente al estadístico. El proceso de ordenación para calcular los intervalos es similar al caso de Matlab ya expuesto.

9. Estimación bayesiana

Thomas Bayes (1702-1761)
Matemático inglés. Fue sacerdote en un pueblo de Kent desde 1731 hasta su muerte. Fue el primero en usar el hoy llamado teorema de Bayes para realizar inferencias. Su trabajo fue publicado póstumamente. Interesado en la astronomía, fue elegido miembro de The Royal Society.

9.1 Introducción

Los métodos de estimación que hemos presentado en los dos capítulos anteriores funcionan muy bien con muestras grandes, pero con muestras pequeñas o medianas no proporcionan siempre respuestas satisfactorias. Consideremos, por ejemplo, que tratamos de estimar la proporción de estudiantes de una universidad que han leído la novela *Rayuela* de J. Cortázar. Supongamos que se pregunta a una muestra aleatoria de 30 estudiantes y el resultado es que ninguno de los miembros de la muestra ha leído esta novela. ¿Qué inferencia podemos hacer? Si aplicamos los métodos estudiados en el capítulo 7, el estimador máximo-verosímil del parámetro de una población binomial es la frecuencia en la muestra, que, en este caso, es cero. La estimación *MV* del número de estudiantes que han leído esta novela es cero. Para cuantificar la precisión de esta estimación nos encontramos que la varianza estimada es cero, o la precisión infinita. ¿Es esto razona-

ble? Claramente no. El problema es que estamos aplicando un método *MV*, que tiene buenas propiedades en muestras grandes pero que puede dar resultados poco satisfactorios en pequeñas e incluso medianas muestras.

Como segundo ejemplo supongamos que tratamos de estimar la edad del más veterano de los estudiantes de una universidad mediante esa misma muestra de tamaño 30. Como desconocemos la distribución de edades, podríamos, por analogía, estimar como valor máximo de la variable edad el mayor valor observado en la muestra. Supongamos que la persona de mayor edad es de 21 años. Esta estimación es intuitivamente muy deficiente: si hay una pequeña proporción de personas mayores en la universidad, probablemente no serán seleccionadas en una muestra pequeña, y el error cometido puede ser muy grande.

Como tercer ejemplo supongamos que sacamos al azar una moneda del bolsillo, la tiramos 10 veces y obtenemos 7 caras y 3 cruces. ¿Quiere esto decir que debemos admitir que la probabilidad de cara en esta moneda es 0,7? Claramente no.

Estos tres ejemplos tienen en común la existencia de cierta información a priori respeto al parámetro que tratamos de estimar, que no se tiene en cuenta en el proceso de inferencia. Ignorar la información inicial, que llamaremos información a priori, que tenemos respecto a un parámetro a estimar no es importante si la muestra es grande, ya que entonces probablemente queremos despreciar nuestra información a priori frente a los datos, pero puede serlo cuando la información a priori sea significativa frente a los datos. La inferencia bayesiana es un procedimiento general para combinar nuestra información a priori con la muestra para obtener una inferencia que tenga en cuenta toda la información existente en el problema.

En el enfoque bayesiano un parámetro no es una constante desconocida, sino una variable aleatoria sobre la que podemos establecer a priori una distribución de probabilidad que refleje nuestro conocimiento del problema. La inferencia respecto a sus posibles valores se obtiene aplicando el cálculo de probabilidades (teorema de Bayes) para combinar la información inicial con la muestral y obtener la distribución del parámetro condicionada a la información disponible. En concreto, suponemos que antes de tomar la muestra se dispone de cierta información respecto al parámetro (o vector de parámetros) que se representa mediante una *distribución inicial o a priori,* $p(\theta)$. En la sección siguiente analizaremos cómo construir estas distribuciones. Después se toma la muestra $\mathbf{X} = (x_1, ..., x_n)$, y la probabilidad de obtener la muestra para cada valor posible del parámetro viene dada por la función de densidad conjunta de las observaciones $f(\mathbf{X}|\theta)$. Observemos que, una vez obtenida la muestra, en esta función los datos son fijos, porque ya han sido observados, mientras que la variable son los parámetros. Por tanto, cuando la muestra se observa, $f(\mathbf{X}|\theta) = \ell(\theta|\mathbf{X})$ es la función de verosimilitud discutida en el capítulo 7. A continuación combinamos según las reglas del cálculo de probabilidades estos dos elementos de información para ob-

tener la *distribución final o a posteriori,* que se obtiene mediante el teorema de Bayes. Llamando $p(\theta|\mathbf{X})$ a la distribución a posteriori, tendremos que

$$p(\theta|\mathbf{X}) = \frac{f(\mathbf{X}|\theta)p(\theta)}{\int f(\mathbf{X}|\theta)p(\theta)d(\theta)} \tag{9.1}$$

La distribución a posteriori contiene toda la información para hacer inferencias respecto al parámetro. Si se desea un estimador puntual, se tomará la media o la moda de dicha distribución; si se desea un intervalo de confianza, se tomará la zona que encierre una probabilidad fijada en dicha distribución. En consecuencia, una vez obtenida la distribución de probabilidad del parámetro, el problema de estimación queda resuelto de manera automática y simple.

Para calcular la distribución a posteriori observemos que el denominador de (9.1) es

$$m(\mathbf{X}) = \int f(\mathbf{X}|\theta)p(\theta)d(\theta)$$

y como función de $\mathbf{X}$ representa la distribución marginal de los datos, con independencia de los valores de los parámetros. Esta distribución se denomina *distribución predictiva* y es una media ponderada de las verosimilitudes $f(\mathbf{X}|\theta)$ por las probabilidades que la distribución a priori asigna a los posibles valores del parámetro. Cuando observamos la muestra el denominador es una constante, la ordenada de la predictiva para los valores de la muestra observada, y el cálculo de (9.1) se simplifica observando que esta constante tiene sólo la función de que la integral de numerador sea la unidad para que el resultado sea una función de densidad, $p(\theta|\mathbf{X})$. Por tanto, llamando k a esta constante y escribiendo:

$$p(\theta|\mathbf{X}) = k\ell(\theta|\mathbf{X})p(\theta) \tag{9.2}$$

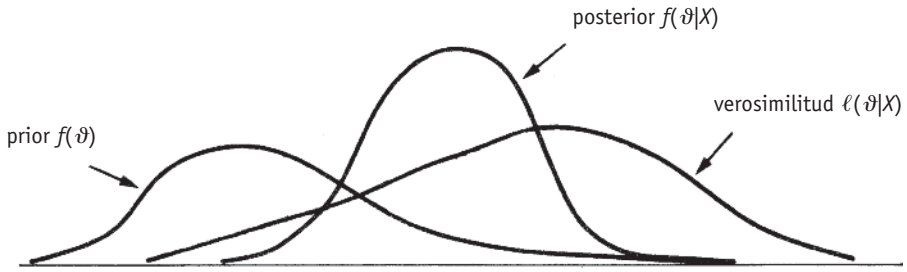
podemos calcular la distribución posterior multiplicando, para cada valor de θ, las ordenadas de $\ell(\theta|\mathbf{X})$ y $p(\theta)$. La constante k es irrelevante para la forma de la posterior, y siempre puede determinarse al final con la condición de que $p(\theta|\mathbf{X})$ sea una función de densidad. El teorema de Bayes puede resumirse en:

$$\text{Posterior} \propto \text{Prior} \times \text{Verosimilitud}$$

donde $\propto$ indica proporcional. La distribución a posteriori es un compromiso entre la prior y la verosimilitud. La figura 9.1 ilustra este cálculo. En el

caso particular de que $p(\theta)$ sea aproximadamente constante sobre el rango de valores en los que la verosimilitud no es nula, se dice que $p(\theta)$ es no informativa, y la posterior vendrá determinada por la función de verosimilitud.

Figura 9.1 Estimación bayesiana



Una ventaja adicional del enfoque bayesiano es su facilidad para procesar información secuencialmente. Supongamos que después de calcular (9.2) observamos una nueva muestra de la misma población **Y**, independiente de la primera. Entonces, la distribución inicial será ahora $p(\theta|\mathbf{X})$ y la distribución final será:

$$p(\theta|\mathbf{XY}) = k\ell(\theta|\mathbf{Y})p(\theta|\mathbf{X})$$

Naturalmente este mismo resultado se obtendría considerando una muestra ampliada (**X**, **Y**) y aplicando el teorema de Bayes sobre dicha muestra, ya que por la independencia de **X** e **Y**:

$$p(\theta|\mathbf{XY}) = k\ell(\theta|\mathbf{XY})p(\theta) = k\ell(\theta|\mathbf{X})p(\theta|\mathbf{Y})p(\theta)$$

La estimación bayesiana proporciona pues un procedimiento automático para expresar el aumento de nuestro conocimiento respecto al parámetro a medida que se recibe información adicional. Éste es uno de sus aspectos más atractivos.

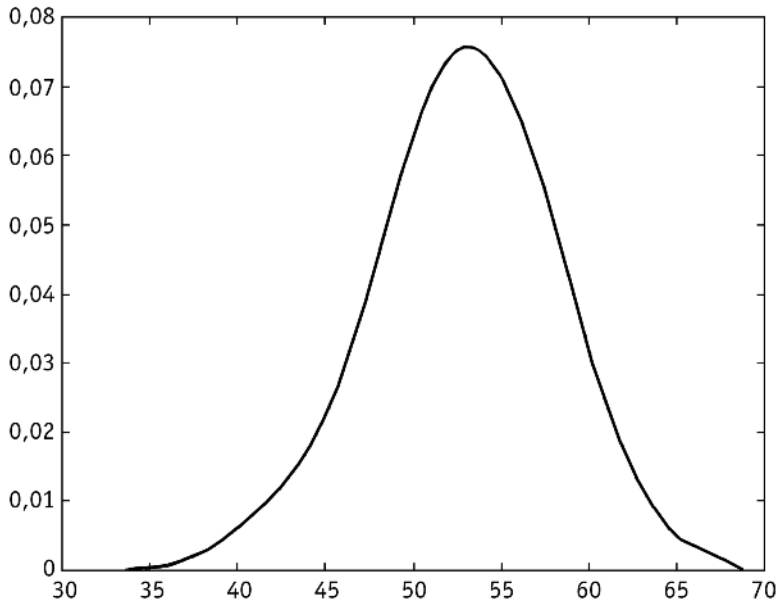
9.2 Distribuciones a priori

La mayor dificultad práctica del enfoque bayesiano es cómo especificar la distribución a priori: normalmente la información de que disponemos es cualitativa y el enfoque bayesiano requiere que establezcamos una distribu-

ción de probabilidad sobre sus valores. Podemos considerar cuatro casos distintos:

1. La distribución a priori proviene de estudios anteriores y se conoce objetivamente. Por ejemplo, supongamos que tratamos de determinar el porcentaje de elementos defectuosos en un proceso. Antes de tomar la muestra conocemos que se hizo un estudio similar hace unos meses, y suponiendo que las condiciones no han cambiado, tomaremos la distribución a posteriori del estudio anterior como distribución a priori del estudio actual. Como segundo ejemplo, supongamos que una empresa esta interesada en conocer el tiempo medio que las personas de una zona dedican a navegar por Internet. Supongamos que conocemos un estudio de esta variable en otra zona de características similares; podemos tomar la distribución a posteriori del estudio realizado como distribución a priori para nuestra zona.

Figura 9.2 Distribución a priori para la edad máxima de un estudiante en una universidad



2. La distribución a priori puede ser importante respecto a la muestral, pero la información existente es subjetiva y no formalizada. Podemos comenzar por decidir el valor más probable del parámetro, que será la moda de la distribución, su rango de valores posibles (o que cubre el 99,9% de la distribución) y si la distribución es o no simé-

trica con relación a la moda. Por ejemplo, en el caso de la moneda, probablemente nuestra opinión a priori sobre la proporción de caras es 0,5, y ésta será la moda de la distribución. Si pensamos que las desviaciones sobre este valor serían debidas a posibles desperfectos por el uso, es razonable suponer una distribución simétrica respecto a 0,5 y con pequeña variabilidad respecto a este valor central. En el caso de los estudiantes universitarios, nuestra estimación a priori dependerá mucho de las características de la universidad: si tiene o no programas para adultos, la importancia del tercer ciclo, etc. Para establecer nuestra opinión podemos fijar el valor más probable, el intervalo central donde debe estar el 50% de la densidad y la forma general de la distribución. Por ejemplo, supongamos que en una universidad sin programas para adultos pero con un amplio programa de tercer ciclo pensamos que el valor más probable (moda) es alrededor de 52 y que estamos seguros de que el estudiante más veterano debe tener más de 35 años y menos de 67. La figura 9.2 presenta una posible distribución a priori para este problema.

3. La información a priori es pequeña con relación a la muestral. Podemos elegir una distribución a priori que refleje globalmente nuestra opinión, en particular la moda a priori y el rango de valores posibles, pero sin preocuparnos mucho del resto de los detalles. En estos casos elegiremos una *distribución conjugada* para el problema, que son distribuciones que facilitan el cálculo de la posterior. Estudiaremos estas distribuciones en la sección siguiente.
4. La información a priori es despreciable frente a la muestral, o no queremos tenerla en cuenta en el proceso de inferencia. En este caso podemos utilizar los métodos clásicos de los capítulos anteriores o utilizar el enfoque bayesiano con una distribución a priori *no informativa* o de *referencia,* que se discuten en la sección siguiente.

9.2.1 Distribuciones conjugadas

El cálculo de la distribución posterior puede ser complicado y requerir métodos numéricos. El problema se simplifica si podemos expresar aproximadamente nuestra información a priori con una distribución que facilite el análisis. Una familia de distribuciones a priori adecuada para este objetivo es aquella que tiene la misma forma que la verosimilitud, de manera que la posterior pueda calcularse fácilmente al pertenecer a la misma familia que la priori. A estas familias se las denomina *conjugadas.*

Una clase $\mathscr{C}$ de distribuciones a priori para un parámetro vectorial θ es conjugada si cuando la prior pertenece a esa clase, $p(\theta) \in \mathscr{C}$ entonces también lo hace la posterior $p(\theta|\mathbf{X}) \in \mathscr{C}$. La distribución conjugada a priori se elige tomando como distribución la verosimilitud, y modificando los valo-

res de las constantes para que la función resultante sea una función de densidad y tenga características coincidentes con nuestra información a priori. Por ejemplo, supongamos que queremos hacer estimar el parámetro θ en un modelo binomial. La verosimilitud es

$$l(\theta) = \theta^r (1 - \theta)^{n-r}$$

La prior conjugada debe ser una función del tipo $\theta^r (1 - \theta)^{n-r}$. El primer paso es modificar las constantes r y n, que determinan la forma de la distribución, para que la función resultante coincida con nuestra opinión. Supongamos que tomamos como distribución a priori:

$$p(\theta) = k\theta^{r_0} (1 - \theta)^{n_0 - r_0}$$

donde k es la constante necesaria para que integre a uno, que dependerá de los parámetros r_0 y n_0. Ésta es la distribución beta que se presentó en el apéndice 5D. La moda de esta distribución es r_0/n_0, y la variabilidad disminuye con n_0. La distribución es simétrica si $r_0 = n_0/2$ y asimétrica en caso contrario. Si a priori el valor más probable para el parámetro es p_0, entonces $p_0 = r_0/n_0$ y podemos elegir n_0 en función de la seguridad que queramos dar a la estimación inicial p_0. De esta manera obtenemos los valores de los parámetros. Veremos un procedimimiento rápido de hacer estas elecciones al estudiar cómo afectan estas estimaciones a la posteriori.

Como segundo ejemplo, supongamos que se trata de estimar la media de una población normal con varianza conocida. La verosimilitud puede escribirse entonces como:

$$l(\theta) = k \exp\left\{-\frac{n}{2\sigma^2}(\bar{x} - \theta)^2\right\}$$

y la prior conjugada será una distribución normal:

$$p(\theta) = k \exp\left\{-\frac{n_0}{2\sigma^2}(\theta - \mu_0)^2\right\}$$

que depende de los dos parámetros μ_0 y n_0. El primero determina la media de la distribución y el segundo la desviación típica. En las secciones siguientes veremos el uso de estas distribuciones conjugadas y métodos rápidos para fijar sus parámetros.

9.2.2 Distribuciones de referencia

Una distribución no informativa o de referencia pretende no modificar la información contenida en la muestra. Intuitivamente, una distribución a priori no informativa para un parámetro de localización es aquella que es localmente uniforme sobre la zona relevante del espacio paramétrico, y escribiremos $p(\theta) = c$, uniforme. Sin embargo, esta elección tiene el problema de que si el vector de parámetros puede tomar cualquier valor real $\int_{-\infty}^{\infty} p(\theta)d\theta = \infty$, y la prior no puede interpretarse como una distribución de probabilidad sino como una herramienta para calcular la posterior. En efecto, si podemos suponer que a priori un parámetro escalar debe estar en el intervalo $(-h, h)$, donde h puede ser muy grande pero es un valor fijo, la distribución a priori $p(\theta) = 1/2h$ es propia, ya que integra a uno. La distribución $p(\theta) = c$ debe pues considerarse como una herramienta simple para obtener la posterior. Estas distribuciones se denominan impropias. En problemas simples trabajar con a prioris impropias no produce problemas, (aunque puede dar lugar a paradojas; véase por ejemplo Bernardo y Smith, 1994), pero en situaciones un poco más complicadas la distribución a posteriori puede no existir si trabajamos con distribuciones impropias.

Las distribuciones constantes están sujetas a una dificultad conceptual adicional: si suponemos que la distribución a priori para un parámetro escalar θ es del tipo $p(\theta) = c$ y hacemos una transformación uno a uno del parámetro $\varphi = g(\theta)$, como

$$p(\varphi) = p(\theta) \left| \frac{d\theta}{d\varphi} \right|$$

si la distribución es constante para el parámetro θ no puede ser constante para el parámetro φ. Por ejemplo, si $p(\theta) = c$, y $\varphi = 1/\theta$, entonces $|d\theta/d\varphi| = \varphi^{-2}$ y $p(\varphi) = c\varphi^{-2}$ que no es uniforme. Nos encontramos con la paradoja de que si no sabemos nada sobre θ y $\theta > 0$, no podemos decir que no sabemos nada (en el sentido de una distribución uniforme) sobre $\log \theta$ o θ^2. Una solución es utilizar las propiedades de invarianza del problema para elegir sobre qué transformación del parámetro es razonable suponer una distribución constante, pero aunque es fácil estar de acuerdo en casos simples, no es inmediato cómo llevar esto a la práctica en general.

Estas dificultades hacen que cuando no se disponga de información relevante, o no queramos utilizarla, lo más simple es trabajar directamente con la verosimilitud, como vimos en el capítulo 7. Así obtenemos de manera simple el mismo resultado que si utilizásemos la distribución de referencia adecuada, ya que, si tenemos muchos datos, la verosimilitud será muy apuntada, y la posterior vendrá determinada por la verosimilitud, al ser esencialmente la priori constante sobre la zona relevante para la inferencia.

9.3 Estimación puntual

Si es necesario elegir un valor único para el parámetro podríamos:

a) Seleccionar el máximo (la moda) de la distribución a posteriori, que es el valor más probable. Cuando la información inicial sea pequeña con relación a la proporcionada por la verosimilitud, la posterior será análoga a la verosimilitud y su moda es el estadístico máximo-verosímil. Por tanto, en este caso el enfoque bayesiano coincide con el *MV*.
b) Definir un criterio de optimalidad y deducir el estimador a partir de él. Esto equivale a definir *una función de pérdida*, $g(\vartheta, \hat{\vartheta})$, que indique la penalización asociada a tomar como estimador $\hat{\vartheta}$ cuando el verdadero valor es ϑ. La función más frecuente es la cuadrática:

$$g(\vartheta, \hat{\vartheta}) = k \cdot (\vartheta - \hat{\vartheta})^2$$

entonces, el criterio de elección será escoger como estimador aquel valor $\hat{\vartheta}$ que haga en promedio la pérdida mínima. La pérdida promedio se denomina *riesgo del estimador* y será:

$$E[g(\vartheta, \hat{\vartheta})] = kE(\vartheta - \hat{\vartheta})^2$$

donde la esperanza se toma respecto a la distribución de ϑ. E1 riesgo será mínimo si $\hat{\vartheta} = E(\vartheta)$, lo que implica que *antes* de observar la muestra la mejor estimación (mínimo riesgo) es la media de la distribución inicial; *después* de observar la muestra, será la media de la distribución final.

Este criterio parece análogo a primera vista al criterio clásico de minimizar el error cuadrático medio. La diferencia entre ambos es que, después de observar la muestra, en el enfoque bayesiano la esperanza se toma respecto a la distribución a posteriori de θ (que resume toda la información disponible), mientras que en el enfoque clásico se toma respecto a la distribución en el muestreo del estimador. En el enfoque clásico este criterio lleva a estimadores centrados de varianza mínima, mientras que en el bayesiano a tomar como estimador la media de la posterior. Para tamaños muestrales grandes la información inicial será escasa con relación a la dada por la muestra, y la posterior será proporcional a la verosimilitud. Asintóticamente la verosimilitud está centrada en el estimador *MV* que, para muestras grandes, es centrado (asintóticamente) y de varianza mínima. Por tanto, para muestras grandes ambos métodos serán similares.

9.4 Estimación de una proporción

Distribución a priori

Una forma simple de expresar la incertidumbre inicial respecto a la proporción de elementos con un atributo en una población (p) es mediante la distribución beta (véase apéndice 5D), cuya función de densidad es:

$$f(p) = kp^{r_0}\,(1-p)^{n_0-r_0} \qquad (0 \le p \le 1) \tag{9.3}$$

y que queda determinada por r_0 y n_0. La moda es r_0/n_0 y la distribución es simétrica si $r_0 = n_0/2$ y asimétrica en otro caso. Una forma de interpretar esta distribución es suponiendo que la información disponible a priori es equivalente a la observación de r_0 elementos con el atributo estudiado en una muestra de n_0 elementos. Cuanto mayor sea n_0, mayor es la cantidad de información disponible, y menor la dispersión de la distribución alrededor de su máximo. Por el contrario, si $r_0 = n_0 = 0$, la función se convierte en la uniforme, y representa una situación sin información inicial relevante.

Cálculo de la posterior

Supongamos que la muestra es de tamaño n y se observa una proporción r/n de elementos con el atributo estudiado. Entonces, la función de verosimilitud es:

$$\ell(p|\mathbf{X}) = kp^{r}(1-p)^{n-r}; \quad 0 \le p \le 1 \tag{9.4}$$

La distribución a posteriori será el producto de (9.4) y (9.3), resultando:

$$f(p|\mathbf{X}) = k'p^{(r+r_0)}\,(1-p)^{(n+n_0)-(r+r_0)} \tag{9.5}$$

y resume toda la incertidumbre respecto al parámetro p.

La moda de esta nueva distribución beta es de nuevo el cociente entre el exponente de p y el primer término en el exponente de $1 - p$, como se comprueba fácilmente calculando el máximo de la función de densidad (9.5). Por tanto, el valor más probable es a priori r_0/n_0, con los datos de la muestra r/n (estimador *MV*) y con toda la información $(r + r_0)/(n + n_0)$. Este estimador puede escribirse:

$$\hat{p} = \frac{r+r_0}{n+n_0} = \left(\frac{n}{n+n_0}\right)\left(\frac{r}{n}\right) + \left(\frac{n_0}{n+n_0}\right)\left(\frac{r_0}{n_0}\right)$$

es decir, llamando $\hat{p}_0$ a la moda de la distribución inicial (r_0/n_0), $\hat{p}_m$ a la proporción observada en la muestra y α al cociente $n/(n + n_0)$:

$$\hat{p} = \alpha\hat{p}_0 + (1 - \alpha)\hat{p}_m \tag{9.6}$$

y la moda de la distribución posterior es una combinación lineal de la inicial y la de la verosimilitud, con ponderaciones iguales a las precisiones relativas de estas estimaciones. En efecto, la precisión de la información inicial depende de n_0 y la de la muestra de n. La fórmula (9.6) resalta el carácter de estimador de compromiso del estimador Bayes y es coherente con los procedimientos clásicos de combinar distintas fuentes de información estudiados en el capítulo 7.

Intervalo de confianza

Para construir un intervalo de confianza de probabilidad $1 - \alpha$ seleccionaremos dos valores en la distribución beta que dejen entre sí el 95%. Si n es grande, la distribución beta se aproxima a la normal, y el intervalo construido a partir de ella estará próximo al intervalo construido con el método clásico.

Ejemplo 9.1

Se conoce que la proporción de un partido *A* en una población está casi con seguridad entre 0,1 y 0,3 y puede representarse adecuadamente por una beta de parámetros $r_0 = 2$, $n_0 = 10$. Se toma una muestra de 30 personas y se obtienen cuatro votantes de *A*. Indicar la distribución a posteriori para el número de votos del partido *A*.

La distribución inicial es:

$$f(p) = kp^2(1 - p)^8$$

Como únicamente nos interesa su forma, podemos prescindir de k o darle cualquier valor conveniente. Por simplicidad tomemos $k = 1000$, con lo que las ordenadas de esta distribución serán (proporcionales) a las dadas en la tabla 9.1.

A continuación calculamos la verosimilitud por el mismo procedimiento:

$$\ell(p) = 10^6p^4\ (1 - p)^{26}$$

donde de nuevo hemos tomado $k = 10^6$ arbitrariamente para obtener números enteros.

La posterior será, multiplicando las ordenadas de ambas funciones,

$$f(p|\mathbf{X}) = k \cdot p^6(1 - p)^{34}$$

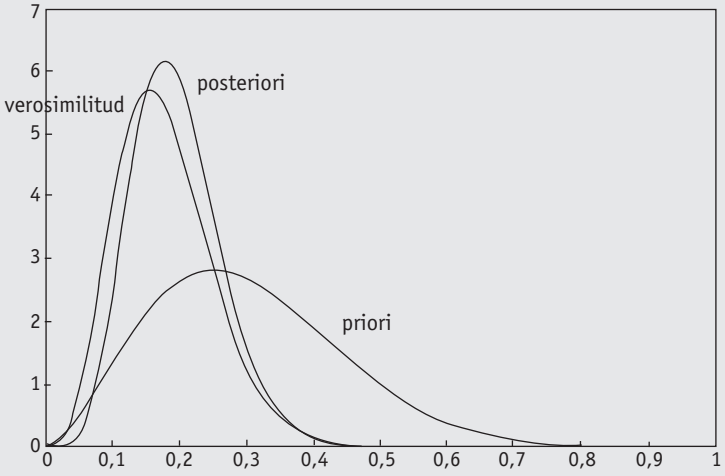
el máximo de esta distribución es en 6/40 = 0,15. Como vemos, este valor es una media ponderada de los valores más probables inicialmente (0,2), y dada la muestra (0,13).

Tabla 9.1 Ordenadas de las tres distribuciones (salvo constantes)

p	0,01	0,05	0,1	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
$f(p)$	0,09	1,66	4,30	6,13	6,71	6,25	5,18	3,90	2,69	1,70	0,98
$\ell(p)$	0	1,65	6,46	7,40	4,84	2,20	0,76	0,20	0,04	0	0
$f(p\|\mathbf{X})$	0	2,7	27,8	45,4	32,5	13,8	3,9	0,8	0,1	0	0

La figura 9.3 representa gráficamente las tres distribuciones.

Figura 9.3 Estimación bayesiana de una proporción



9.5 Estimación de la media en poblaciones normales

Varianza conocida

Supongamos que se desea estimar la media de una población normal con varianza conocida y que la información inicial respecto a μ se traduce en una distribución a priori $N(\mu_0;\ \sigma_0)$. Utilizando los resultados del ejemplo 7.10, la verosimilitud es:

$$\ell(\mu|\mathbf{X}) = k \exp\left\{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2\right\}$$

y la posterior será:

$$f(\mu|\mathbf{X}) = k' \exp\left\{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2 - \frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right\}$$

que depende de la muestra únicamente a través del valor de $\bar{x}$. Su exponente puede escribirse (véase ejercicio 9.6):

$$\frac{n}{\sigma^2}(\bar{x}-\mu)^2+\frac{1}{\sigma_0^2}(\mu-\mu_0)^2=\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)(\mu-\mu_p)^2+\frac{n}{n\,\sigma_0^2+\sigma^2}(\bar{x}-\mu_0)^2 \qquad (9.7)$$

donde:

$$\mu_p = \frac{\dfrac{n}{\sigma^2}\bar{x} + \dfrac{1}{\sigma_0^2}\mu_0}{\dfrac{n}{\sigma^2}+\dfrac{1}{\sigma_0^2}} \qquad (9.8)$$

La descomposición (9.7) equivale a:

$$f(\bar{x}|\mu)f(\mu) = f(\mu|\bar{x})f(\bar{x})$$

y en el segundo miembro aparecen los exponentes de la posterior, $f(\mu|\bar{x})$, y de la predictiva, $f(\bar{x})$, que no depende de μ. Llamando $p_{\bar{x}}$ y p_0 a la precisión muestral y a priori, la distribución posterior es una normal con parámetros:

$$\mu_p = \frac{p_{\bar{x}}\,\bar{x} + p_0\mu_0}{p_{\bar{x}} + p_0} \qquad (9.9)$$

$$\frac{1}{\sigma_p^2} = p_p = p_{\bar{x}} + p_0 \qquad (9.10)$$

es decir, la media posterior es una combinación lineal de la prior y la muestral, con pesos que dependen de la precisión relativa. La precisión final es la suma de la inicial y la verosimilitud.

Otra forma ilustrativa de escribir estos resultados es definiendo:

$$n_0 = \frac{\sigma^2}{\sigma_0^2}$$

como el cociente entre la varianza de la población y la de la distribución a priori. Sustituyendo σ_0^2 en (9.9) y (9.10) podemos escribir:

$$\mu_p = \frac{n\bar{x} + n_0\mu_0}{n + n_0} \tag{9.11}$$

$$\sigma_p = \frac{\sigma}{\sqrt{n + n_0}} \tag{9.12}$$

con lo que queda de manifiesto el carácter de información adicional de la muestra. Observemos que, de nuevo, las ecuaciones (9.9) o (9.11) coinciden con el procedimiento clásico de combinar distintas fuentes de información independientes.

Cuando la información inicial sea vaga respecto a la muestral, σ_0^2 será mucho mayor que σ_n^2 y la distribución a posteriori vendrá determinada por la verosimilitud.

Varianza desconocida, muestras grandes

Cuando σ^2 es desconocida, la estimación bayesiana requiere establecer una distribución a priori sobre ambos parámetros. Se demuestra que la distribución final marginal para el parámetro μ es una t generalizada con media μ_p, dada por:

$$\mu_p = \frac{n\bar{x} + n_0\mu_0}{n + n_0} \tag{9.13}$$

donde $n_0 = \hat{s}_0^2/\sigma_0^2$, y factor de escala:

$$\sigma_p = \frac{s}{\sqrt{n + n_0}} \tag{9.14}$$

Para tamaños muestrales grandes ($n > 30$) podemos aproximar la t por la normal y utilizar como distribución final una normal con parámetros (9.13) y (9.14).

Ejemplo 9.2

Una empresa realiza un estudio de mercado para decidir si lanzar o no un nuevo producto. La variable clave son las ventas mensuales medias por punto de venta, sobre la que existe bastante incertidumbre, representada por una distribución inicial $N(500; 30)$. Se realiza un test de 25 puntos de venta, obteniendo unas ventas medias en ellos de 535 unidades con desviación típica de 20 unidades. ¿Qué podemos concluir?

La verosimilitud para μ, tomando $\hat{s} = 20$ como estimador de σ, será aproximadamente normal con media 535 unidades y desviación típica $20/\sqrt{25} = 4$. Por tanto, la media posterior será:

$$\mu_p = \frac{1/16}{1/16 + 1/900}\, 535 + \frac{1/900}{1/16 + 1/900}\, 500 =$$

$$= (0{,}98)\ 535 + (0{,}02)\ 500 = 534{,}3$$

y la precisión:

$$p = \frac{1}{900} + \frac{1}{16} = 0{,}0636$$

que implica:

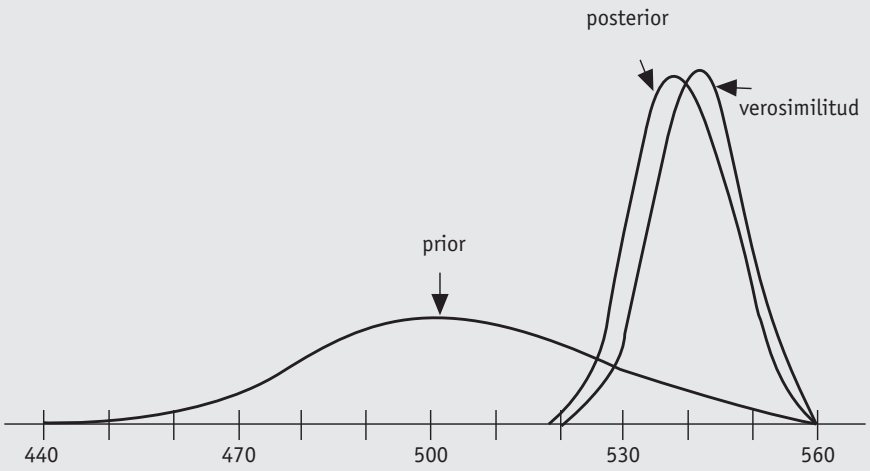
$$\sigma_p = \frac{1}{\sqrt{0{,}0636}} = 3{,}96$$

La figura 9.4 presenta la distribución a priori y la posterior, que es casi análoga a la verosimilitud. Un intervalo de probabilidad 95% para μ se determina fácilmente como:

$$534{,}3 \pm 2{,}06\ (3{,}96)$$

donde 2,06 es el percentil (0,975) de la distribución t con 24 grados de libertad.

Figura 9.4



El intervalo del 95% clásico, sin tener en cuenta la distribución inicial, es:

$$535 \pm 2{,}06 \cdot 4$$

Se observa que el intervalo bayesiano es algo más corto, resultado lógico ya que utiliza mayor información.

El resultado es análogo si aproximamos la distribución final con la normal en lugar de la t.

9.6 Comparación con los métodos clásicos

Las diferencias prácticas entre el enfoque clásico basado en la función de verosimilitud y el bayesiano pueden ser importantes en muestras pequeñas y son irrelevantes en muchas grandes. La ventaja principal del enfoque bayesiano es su simplicidad conceptual, su generalidad y la capacidad de incluir información adicional al proceso de inferencia. En contrapartida exige una estructura formal más rígida que, cuando el tamaño muestral es

grande, puede no aportar ventajas adicionales al método de máxima verosimilitud.

Los estimadores puntuales obtenidos por máxima verosimilitud coinciden, en muestras grandes, con la moda de la distribución posterior, que será próxima a la media por la normalidad asintótica de esta distribución. Las distribuciones a posteriori son análogas a las distribuciones de confianza que proporcionan los intervalos de confianza en el método clásico.

El cálculo de las distribuciones a posteriori puede ser complejo, pero en muchos problemas es posible obtenerlas fácilmente con un ordenador generando muestras de la distribución posterior a partir de la prior y de la verosimilitud. Existe una variedad de métodos para realizar esta simulación que se conocen bajo el nombre común de métodos de Monte Carlo con cadenas de Markov (o métodos MC^2). Uno de los métodos más utilizados es el muestreo de Gibbs, que permite obtener muestras de una distribución conjunta si se conocen las distribuciones condicionadas. En la sección siguiente el lector puede encontrar referencias de este método.

Ejercicios 9

9.1. Para estudiar el gasto medio semanal de un estudiante universitario se toma una muestra aleatoria simple de 25 estudiantes obteniendo $\bar{x} = 18$ euros; $\hat{s} = 1{,}8$ euros. Construir un intervalo donde se encuentre la media con probabilidad 95% si a priori $\mu \sim N(15; 3)$.

9.2. Repita los cálculos del ejercicio anterior incluyendo su propia distribución a priori.

9.3. ¿Cuál tendría que ser la desviación típica de la distribución inicial para que la media a posteriori fuese 1.700?

9.4. La proporción de artículos defectuosos en un lote es o bien 0,05 o bien 0,01. A priori se supone que $P(0{,}05) = 0{,}8$; $P(0{,}01) = 0{,}2$. Se toma una muestra de tres elementos y los tres son buenos. ¿Cuál es la distribución final?

9.5. Se selecciona una observación x de una variable uniforme $(\vartheta - \frac{1}{2};\ \vartheta + \frac{1}{2})$ y la distribución inicial de ϑ es uniforme en (5,15). Si el valor observado es 10, ¿cuál es la distribución final?

9.6. Demostrar la igualdad $A(x-a)^2 + B(x-b)^2 = (A+B)(x-c)^2 + \dfrac{AB}{A+B}(a-b)^2$ con $c = (Aa + Bb)/(A + B)$ y utilizarla para demostrar la ecuación (9.7).

9.7 Resumen del capítulo y consejos de cálculo

El enfoque bayesiano de inferencia permite incorporar información adicional a la muestra al proceso de inferencia. Esta incoporación se realiza siguiendo las reglas de las probabilidades mediante el teorema de Bayes. Para ello es imprescindible establecer una distribución a priori sobre los parámetros. Si se quiere dejar a los datos hablar por sí mismos, la distribución a priori se toma como no informativa o de referencia. Para facilitar el cálculo de la posterior pueden utilizarse distribuciones a priori conjugadas, que facilitan el cálculo de la posterior.

La distribución a posteriori resume toda la información para la inferencia. El cuadro 9.1 recoge los resultados principales obtenidos en este capítulo.

Los programas habituales no proporcionan directamente estimadores bayesianos. Sin embargo, es fácil programarlos para realizarlos, y en la red existen muchos programas con esta orientación. Por ejemplo, programas en Matlab y Minitab para el cálculo bayesiano se encuentran en http://www.math.bgsu.edu/~albert/. Un programa para iniciarse en la inferencia bayesiana se encuentra en http://www.shef.ac.uk/~stlao/1b.html. Statlib, que contiene muchos programas de estadística, incluye también paquetes bayesianos (http://lib.stat.cmu.edu/).

Cuadro 9.1 Estimación bayesiana

Población	Parámetro	Distribución inicial	Distribución final
$f(x \mid \vartheta)$	ϑ	$f(\vartheta)$	$f(\vartheta \mid \mathbf{x}) = kf(\mathbf{x} \mid \vartheta)f(\vartheta)$
Binomial $k_1 p^r (1-p)^{n-r}$	p	$k_0 p^{r_0} (1-p)^{n_0-r_0}$ Beta (r_0, n_0) $\hat{p}_0 = r_0/n_0$	$k_2 p^{r-r_0} (1-p)^{n+n_0-(r-r_0)}$ Beta $(r_0 + r; n + n_0)$ $\hat{p} = \alpha\hat{p}_0 + (1-\alpha)r/n$ $\alpha = n/(n + n_0)$
Normal (μ, σ)	μ σ conocido	$N(\mu_0, \sigma_0)$ $n_0 = \sigma^2/\sigma_0^2$	$N(\mu_p, \sigma_p)$ $\mu_p = \alpha\mu_0 + (1-\alpha)\bar{x}$ $\alpha = n_0/(n + n_0)$ $\sigma_p = \sigma/(n + n_0)^{1/2}$
Normal (μ, σ)	μ muestras grandes	$N(\mu_0, \sigma_0)$ $n_0 = \sigma^2/\sigma_0^2$	$N(\mu_p, \sigma_p)$ $\mu_p = \alpha\mu_0 + (1-\alpha)\bar{x}$ $\sigma_p = \hat{s}/(n + n_0)^{1/2}$

9.8 Lecturas recomendadas

Dos libros recomendables para iniciarse en la inferencia bayesiana son Lee (2004) y Berry (1996). Lindley (1970), Winkler (2003) y Antelman (1997) son también muy claros y fáciles de leer. Libros excelentes y con tratamientos modernos aunque más avanzados son Gelman *et al.* (2003) y O'Hagan (2004). Referencias más extensas son Bernardo y Smith (2000), que presentan un tratamiento muy completo de los fundamentos, Berger (1993) y Robert (2007), que ponen énfasis en el enfoque decisional, Press (2002), con orientación a los métodos multivariantes, y Box y Tiao (1992). Una comparación clara de las distintas filosofías de inferencia aparece en Barnett (1999) y De Groot (1988), y a un nivel más detallado y matemático, en Cox y Hinkley (1979). Para los métodos bayesianos de cálculo intensivo el lector interesado puede acudir a Robert y Casella (2005), Carlin y Louis (2000) y Gaberman y Lopes (2006).

10. Contraste de hipótesis

Egon Pearson (1895-1980)
Científico británico hijo de K. Pearson. Creador con Neyman de la teoría de contrastes de hipótesis. Cuando K. Pearson se retiró, su cátedra en University College in London se dividió en estadística, para E. Pearson, y eugenesia, para Fisher, y entre ambos hubo desde entonces amplias discrepancias. E. Pearson ha hecho también importantes contribuciones a la historia de la estadística.

10.1 Introducción

Un principio general de la investigación científica es escoger siempre la hipótesis más simple capaz de explicar la realidad observada. La razón es que una hipótesis simple es más fácil de contrastar empíricamente y descubrir sus deficiencias, lo que permite aprender de los datos con mayor rapidez y seguridad.

Este principio justifica que muchas investigaciones estadísticas tengan por objeto contrastar una hipótesis simplificadora del tipo: una población es idéntica a otra de referencia; dos o más poblaciones son iguales entre sí. Por ejemplo, se conoce que la vida media de los elementos resultantes de un proceso de fabricación es 5.000 horas, se introducen cambios en el proceso y se desea contrastar que la vida media no ha variado. Como segundo ejemplo, se desea saber si la remuneración media obtenida en un trabajo

análogo por personas de igual cualificación profesional es la misma (no depende del sexo). Como tercero, ocurre un cambio legal que puede afectar al precio medio de las viviendas en una zona y se contrasta que la ley no ha tenido efectos y que los precios medios (descontados otros factores) antes y después de la ley son análogos.

Una hipótesis se contrasta comparando sus predicciones con la realidad: si coinciden, dentro del margen de error admisible, mantendremos la hipótesis; en caso contrario, la rechazaremos, y buscaremos nuevas hipótesis capaces de explicar los datos observados. Este proceso iterativo es consustancial al avance de cualquier disciplina científica. La metodología utilizada cuando existe incertidumbre, y las predicciones generadas por la hipótesis tengan que hacerse en probabilidad, es la teoría estadística de contraste de hipótesis, que expondremos a continuación.

Un problema aparentemente muy distinto, pero que está relacionado con el anterior, es decidir entre cursos alternativos de acción en condiciones de incertidumbre. Por ejemplo, decidir si revisar o no un proceso de fabricación que puede estar desajustado, o lanzar o no un nuevo producto al mercado, o comprar o no una nueva maquinaria. La metodología para analizar estos problemas es la teoría de la decisión, que incluye como caso particular la teoría de contraste de hipótesis. El capítulo 11 presenta una introducción a esta teoría.

Contrastar una hipótesis requiere comparar las predicciones que se derivan de ella con los datos observados. Cuando exista variabilidad, o errores de medida, esta contrastación debe hacerse estadísticamente. Como ejemplo, consideremos un proceso de fabricación que, en condiciones normales, produce elementos cuya vida distribuye normalmente con media 5.000 horas y desviación típica 100 horas. Se introducen ciertos cambios en el proceso que pueden afectar a la media pero no a la variabilidad. Para contrastar si estos cambios han producido efectos, se toma una muestra de cuatro elementos cuyas vidas resultan ser 5.010 h., 4.750 h., 4.826 h. y 4.953 h. ¿Hay evidencia de un efecto sobre la media?

Las dos hipótesis posibles en este caso son H_0: no hay efectos y μ = 5.000 h.; H_1 hay efectos y $\mu \neq 5.000$ h. Tomaremos como hipótesis básica la primera, ya que es más simple, y adoptaremos la postura de admitirla a no ser que se demuestre lo contrario. (Si pensamos a priori que el cambio debe afectar mucho, lo anterior no sería razonable, ya que H_0 sería muy inverosímil.) Entonces, si H_0 es cierta, podemos predecir que la media muestral $\bar{x}$ de una muestra aleatoria simple de tamaño cuatro equivale a una extracción al azar de una distribución normal con media 5.000 h. y desviación típica 50 = $100/\sqrt{4}$ horas. En consecuencia, podemos prever que:

$$|\bar{x} - 5.000| \leq 1{,}96 \cdot 50 = 98$$

es decir, no esperamos que la media muestral se separe de la media poblacional más de 98 horas con probabilidad 95%. En otros términos, la media muestral debe estar, con probabilidad 95%, en el intervalo:

$$4.902 \leq \bar{x} \leq 5.098$$

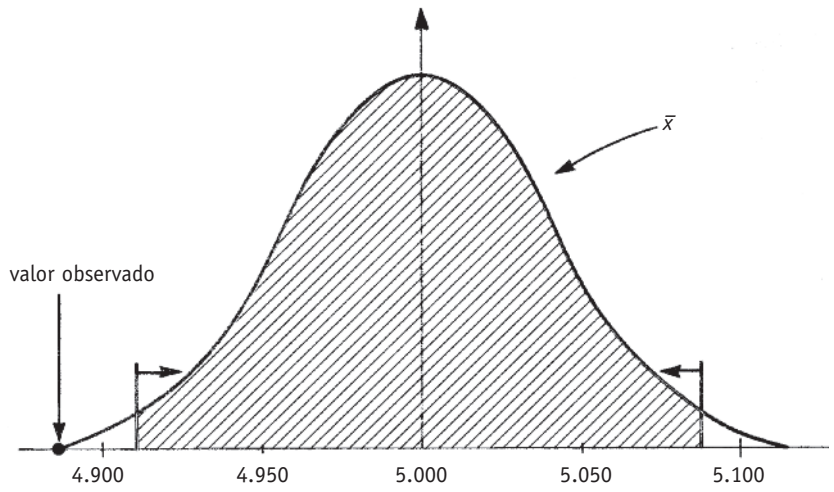
Como la media muestral observada es 4.884,75, la hipótesis H_0 ha sido incapaz de prever, con probabilidad 95%, lo observado. La figura 10.1 muestra la distribución de $\bar{x}$ cuando H_0 es cierta, el intervalo construido y el valor observado. Se observa que este dato es muy improbable cuando H_0 es cierta. Ante este hecho caben dos opciones:

1) continuar aceptando H_0, y atribuir la discrepancia al azar;
2) rechazar H_0 y concluir que se ha producido un cambio. Entonces $\bar{x}$ vendría de una distribución con media menor de 5.000, lo que explicaría el hecho observado.

Para decidir entre ambas alternativas es conveniente indicar *antes* de observar la muestra qué grado de evidencia es necesario para rechazar H_0. Cuanto más convencidos estemos de que H_0 es cierta, más evidencia hará falta para rechazarla con los datos muestrales.

En las secciones siguientes desarrollamos estas ideas.

Figura 10.1 Contraste para la media de una población



10.2 Tipos de hipótesis

Llamaremos hipótesis estadística a una suposición que determina, parcial o totalmente, la distribución de probabilidad de una o varias variables aleatorias. Estas hipótesis pueden clasificarse, según que:

1) Especifiquen un valor concreto o un intervalo de valores para los parámetros de una variable.
2) Establezcan la igualdad de las distribuciones de dos o más variables (poblaciones).
3) Determinen la forma de la distribución de la variable.

Un ejemplo del primer tipo es que la media de una variable es 10; del segundo, que las medias de dos poblaciones normales con igual varianza son idénticas; del tercero, que la distribución de una variable es normal. Aunque la metodología para realizar el contraste es análoga en los tres casos, es importante distinguir entre ellos porque:

1) la contrastación de una hipótesis respecto a un parámetro está muy relacionada con la construcción de intervalos de confianza, y tiene frecuentemente una respuesta satisfactoria en términos de estimación;
2) la comparación de dos o más poblaciones requiere en general un *diseño experimental* que asegure la homogeneidad de las comparaciones;
3) un contraste sobre la forma de la distribución es un contraste no paramétrico que debe realizarse dentro de la fase de validación del modelo que estudiaremos en el capítulo 12.

En este capítulo estudiaremos principalmente las hipótesis del primer tipo y comenzaremos las del segundo, que se desarrollan en el segundo tomo dentro del bloque de diseños de experimentos. Las hipótesis del tercer tipo se estudiarán en el capítulo 12.

Llamaremos *hipótesis simples* a aquellas que especifican un único valor para el parámetro (por ejemplo $\vartheta = \vartheta_0$) e *hipótesis compuestas* a las que especifican un intervalo de valores (ejemplo: $\vartheta > \vartheta_0$; $a \leq \vartheta \leq b$).

10.2.1 Hipótesis nula

Llamaremos hipótesis nula, H_0, a la hipótesis que se contrasta. El nombre de «nula» proviene de que H_0 representa la hipótesis que mantendremos a no ser que los datos indiquen su falsedad, y debe entenderse, por tanto, en el sentido de «neutra». La hipótesis H_0 nunca se considera probada, aunque

puede ser rechazada por los datos. Por ejemplo, la hipótesis de que todos los elementos de las poblaciones A y B son idénticos puede ser rechazada encontrando elementos de A y B distintos, pero no puede ser «demostrada» más que estudiando todos los elementos de ambas poblaciones, tarea que puede ser imposible.

Análogamente, la hipótesis de que dos poblaciones tienen la misma media puede ser rechazada fácilmente cuando ambas difieran mucho, analizando muestras suficientemente grandes de ambas poblaciones, pero no puede ser «demostrada» mediante muestreo (es posible que las medias difieran en δ, siendo δ un valor pequeño imperceptible en el muestreo).

La hipótesis H_0 se elige normalmente de acuerdo con el *principio de simplicidad científica*, que podríamos resumir diciendo que solamente debemos abandonar un modelo simple a favor de otro más complejo cuando la evidencia a favor de este último sea fuerte.

En consecuencia, en el primer tipo de contrastes respecto a los parámetros de una distribución, la hipótesis nula suele ser que el parámetro (o vector de parámetros) es igual a un valor concreto que se toma como referencia. Cuando comparamos poblaciones, H_0 es siempre que las poblaciones son iguales. Cuando investigamos la forma de la distribución H_0 suele ser que los datos son una muestra homogénea de una población simple (normal, Poisson, etc.).

La metodología que vamos a exponer sigue el principio de simplicidad, y tiende a primar a H_0. Si el problema corresponde a elegir entre dos hipótesis que, a priori, se consideran equivalentes, esta metodología es muy discutible, y si las consecuencias de los errores pueden cuantificarse, un enfoque más adecuado es la teoría de la decisión, que estudiaremos en el capítulo 11.

10.2.2 Hipótesis alternativa

Si rechazamos H_0 estamos implícitamente aceptando una hipótesis alternativa, H_1. Suponiendo que H_0 es simple, del tipo $\vartheta = \vartheta_0$, los casos más importantes de hipótesis alternativas son:

a) desconocemos en qué dirección puede ser falsa H_0, y especificamos H_1: $\vartheta \neq \vartheta_0$; decimos entonces que el *contraste es bilateral;*
b) conocemos que si $\vartheta \neq \vartheta_0$ forzosamente $\vartheta > \vartheta_0$ (o bien $\vartheta < \vartheta_0$). Por ejemplo, se introducen cambios en un proceso que, si afectan, aumentan la vida media de los elementos fabricados pero no pueden disminuirla. Tenemos entonces un *contraste unilateral.*

10.3 Metodología del contraste

La metodología actual de contraste de hipótesis es el resultado de los trabajos de R. A. Fisher, J. Neyman y E. S. Pearson entre 1920 y 1933. Su lógica es similar a la de un juicio penal, donde debe decidirse si el acusado es inocente o culpable. Entonces, la hipótesis nula es que el acusado es inocente, y el juicio consiste en aportar evidencia suficiente para rechazar esta hipótesis de inocencia más allá de cualquier duda razonable. Análogamente, un contraste de hipótesis analiza si los datos observados permiten rechazar la hipótesis nula, comprobando si éstos tienen una probabilidad de aparecer lo suficientemente pequeña cuando es cierta la hipótesis nula. En síntesis, las etapas del contraste son:

1) Definir la hipótesis nula a contrastar, H_0, y la hipótesis alternativa, H_1. Los dos casos más importantes de contrastes paramétricos son H_0 simple ($\vartheta = \vartheta_0$) *y* H_1 bilateral ($\vartheta \neq \vartheta_0$), y H_0 compuesta ($\vartheta \leq \vartheta_0$) y H_1 unilateral ($\vartheta > \vartheta_0$). Este segundo caso equivale al contraste simple $\vartheta = \vartheta_0$ frente al unilateral $\vartheta > \vartheta_0$, por lo que en adelante supondremos que H_0 es del tipo $\vartheta = \vartheta_0$.
2) Definir una medida de discrepancia entre los datos muestrales, **X** y la hipótesis H_0. Para contrastes paramétricos la discrepancia puede expresarse como una función del valor del parámetro especificado por H_0 y el valor estimado en la muestra, $\hat{\vartheta}$:

$$d(\vartheta_0 \ ; \ \hat{\vartheta})$$

   La medida de discrepancia debe tener una distribución conocida cuando H_0 sea cierta. De esta manera podemos decir que una discrepancia es grande cuando tiene una probabilidad muy pequeña de ocurrir cuando H_0 es cierta, y pequeña cuando es esperable si H_0 es cierta.
3) Decidir qué discrepancias consideramos inadmisibles con H_0, es decir, a partir de qué valor la diferencia entre $\hat{\vartheta}$ y ϑ_0 es demasiado grande para poder atribuirse al azar.
4) Tomar la muestra, calcular el estimador $\hat{\vartheta}$ y la discrepancia *d*. Si ésta es pequeña, aceptar H_0; si es demasiado grande, rechazar H_0 y aceptar H_1.

   Definir un contraste de significación requiere, por tanto:

a) Una medida de discrepancia.
b) Una regla para juzgar qué discrepancias son «demasiado» grandes.

10.3.1 Medidas de discrepancia

La medida de discrepancia depende de la hipótesis alternativa. En contrastes bilaterales el signo de la desviación entre $\hat{\vartheta}$ y ϑ_0 es irrelevante, por lo que es natural considerar medidas de discrepancia del tipo:

$$d_1 = \left| \frac{\vartheta_0 - \hat{\vartheta}_{MV}}{\hat{\sigma}_{MV}} \right|$$

donde $\hat{\vartheta}_{MV}$ es el estimador *MV* de ϑ y $\hat{\sigma}_{MV}$ su desviación típica. Entonces, d_1 tiene una distribución conocida, ya que, aproximadamente, en muestras grandes:

$$P(d_1 \leq a|H_0) = P(|z| \leq a) = P(-a \leq z \leq a)$$

donde z es $N(0, 1)$.

Cuando el contraste es unilateral, el signo de la desviación es importante. Por ejemplo, si H_1 es $\vartheta > \vartheta_0$, la discrepancia con $H_0(\vartheta = \vartheta_0)$ será tanto mayor cuanto mayor sea la diferencia entre el estimador y ϑ_0, lo que conduce a medidas del tipo:

$$d_2 = \begin{cases} 0 & \text{si } \hat{\vartheta}_{MV} \leq \vartheta_0 \\ \dfrac{\hat{\vartheta}_{MV} - \vartheta_0}{\hat{\sigma}_{MV}} & \text{si } \hat{\vartheta}_{MV} \geq \vartheta_0 \end{cases}$$

donde para d_2 positiva (que es la zona de interés) las probabilidades se calculan de nuevo con la normal estándar.

En el capítulo anterior vimos que el estimador *MV* conducía a buenos intervalos, por lo que es razonable esperar que conduzca a buenos contrastes. Como veremos a continuación, esta intuición es correcta.

10.3.2 Nivel de significación y región de rechazo

El método tradicional de realizar un contraste es dividir el rango de discrepancias que puede observarse cuando H_0 es cierta en dos regiones: una región de aceptación de H_0 y otra de rechazo. Se consideran discrepancias «demasiado grandes» las que tienen una probabilidad pequeña α (normalmente 0,05, 0,01 o 0,001) de ocurrir si H_0 es cierta. Si rechazamos H_0 cuando ocurre una discrepancia de probabilidad α, este número puede interpretarse como la probabilidad que estamos dispuestos a asumir de rechazar H_0 cuando es cierta. Escribiremos:

nivel de significación (α) = P(rechazar $H_0|H_0$ es cierta)

Fijado α, la región de rechazo se determina a partir de la distribución de $d(\hat{\vartheta}, \vartheta_0)$ cuando H_0 es cierta. Como esta distribución es conocida, elegiremos d_c de manera que:

$$P(d > d_C|H_0 \text{ es cierta}) = \alpha$$

Por tanto, discrepancias mayores que d_C tienen una probabilidad de ocurrir menor que α si H_0 es cierta. La región de rechazo será:

$$d > d_C$$

y la región de aceptación o de no rechazo de H_0 será la complementaria:

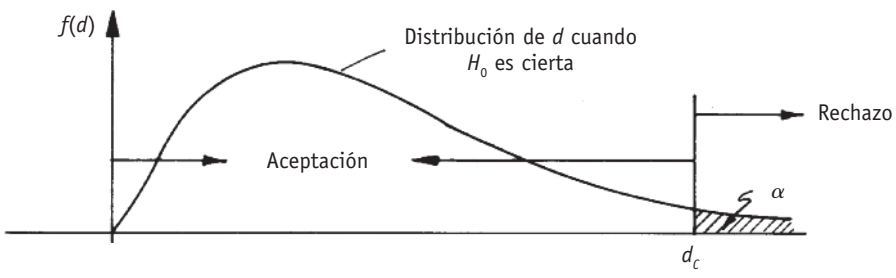
$$d \leq d_C$$

La figura 10.2 muestra gráficamente este método. Si la discrepancia observada en la muestra, d, cae en la región de rechazo, rechazaremos H_0; en caso contrario, la aceptaremos.

Diferencias estadísticamente significativas

Cuando la discrepancia observada en la muestra pertenece a la región de rechazo, se dice que se ha producido una *diferencia significativa*, y se rechaza la hipótesis H_0.

Figura 10.2 Nivel de significación de un test



Esta terminología hace referencia a un concepto estadístico que puede tener poca relación con la significatividad práctica. Por ejemplo, se trata de contrastar si la producción media de una máquina es μ_0; si tomamos una muestra muy grande, es bastante probable que observemos una diferencia significativa y rechacemos que la media es μ_0. Sin embargo, la conclusión puede ser que la media es

$$\mu = \mu_0 + 0{,}00001$$

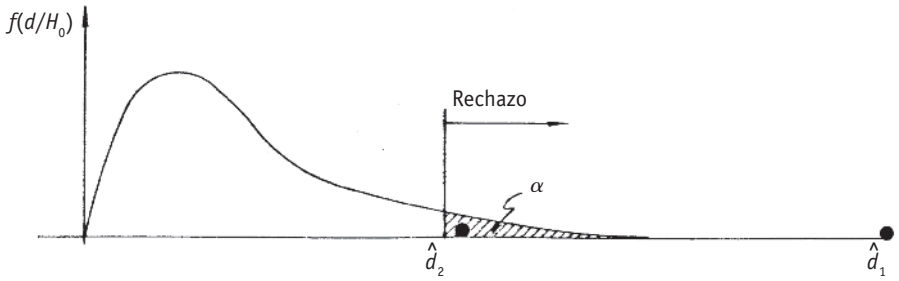
y la diferencia entre μ y μ_0 puede ser perfectamente irrelevante en la práctica.

Críticas a la selección del nivel de significación

El procedimiento de selección de una región de rechazo mediante el nivel de significación está sujeto a tres críticas principales:

1) El resultado del test puede depender mucho del valor de α, que es arbitrario, siendo posible rechazar H_0 con $\alpha = 0{,}05$ y aceptarla con $\alpha = 0{,}04$.
2) Dar sólo el resultado del test no permite diferenciar el grado de evidencia que la muestra indica a favor o en contra de H_0. En la figura 10.3 tanto $\hat{d}_1$ como $\hat{d}_2$ conducen a rechazar H_0, aunque con evidencia muy distinta.

Figura 10.3 Dos muestras proporcionan distinta evidencia



3) Si H_0 especifica el valor de un parámetro y el test conduce a rechazarlo, conviene indicar su estimación a la vista de los datos, para distinguir la significatividad estadística de la práctica.

Un procedimiento para hacer frente a las dos primeras críticas es utilizar en lugar del nivel de significación (α) el nivel crítico de un test, p, que definiremos a continuación.

10.3.3 El nivel crítico *p*

Se define el *nivel crítico p* del contraste como la probabilidad de obtener una discrepancia mayor o igual que la observada en la muestra, cuando H_0 es cierta. Es decir, llamando $\hat{d}$ al valor observado.

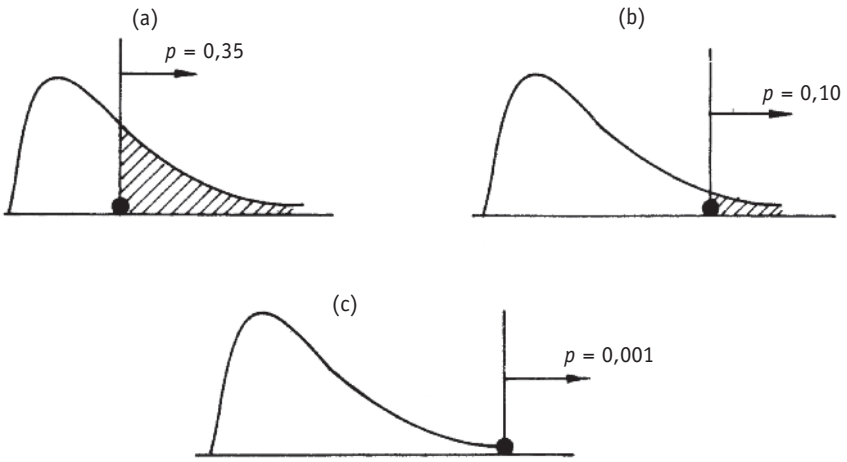
$$p = \text{Prob}(d \geq \hat{d}/H_0)$$

Por tanto, el valor de p *no* se fija a priori, sino que se determina a partir de la muestra.

En la figura 10.3, si $\alpha = 0{,}05$, el valor crítico p es del orden de 0,04 para $\hat{d}_2$ y de 0,0001 para $\hat{d}_1$. Cuanto menor sea p, menor es la probabilidad de aparición de una discrepancia como la observada, y menor la credibilidad de H_0.

Cuando el nivel crítico p es, aproximadamente, mayor que 0,25 (figura 10.4[a]), no existe claramente en la muestra evidencia para rechazar la hipótesis. Cuando este valor está entre 0,2 y 0,01 (caso 10.4[b]), el rechazo o no de la hipótesis dependerá de nuestra opinión a priori y de las consecuencias prácticas de aceptar y rechazar H_0. Finalmente, cuando el nivel crítico es menor que 0,01, rechazaremos, en general, H_0.

Figura 10.4 Nivel crítico de un test y sus consecuencias



La aceptación o rechazo de H_0 depende de tres componentes:

1) La opinión a priori que tengamos de su validez.
2) Las consecuencias de equivocarnos.
3) La evidencia aportada por la muestra.

El nivel de significación se fija en función de los dos primeros, mientras que el nivel crítico permite poner de manifiesto el tercero, dejando al investigador que elabore sus propias conclusiones. Si las consecuencias pueden cuantificarse, un enfoque sistemático del problema es la teoría de decisión, como veremos en el capítulo 11.

10.3.4 Potencia de un contraste

Definir un contraste equivale a definir una medida de discrepancias, α, y una región de rechazo $d \geq d_C$. Hemos visto que d_C se obtiene a partir del nivel de significación, que es la probabilidad de cometer el llamado *error tipo I*: rechazar H_0 cuando es cierta.

Sin embargo, existe otro posible error: aceptar H_0 cuando es falsa. Este error se denomina error tipo II y, definido un contraste, su magnitud depende del verdadero valor del parámetro. Llamaremos:

$$\beta(\vartheta) = P(\text{aceptar } H_0 | \vartheta_0)$$

función o curva característica del contraste. Para $\vartheta = \vartheta_0$, se verifica:

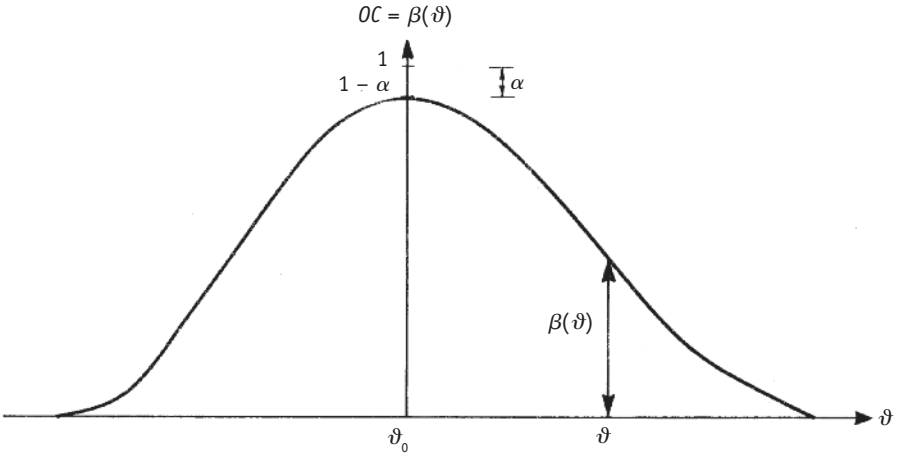
$$\beta(\vartheta_0) = P(\text{aceptar } H_0 | \vartheta_0) = 1 - \alpha$$

mientras que para cualquier otro valor, $\beta(\vartheta)$ proporciona la probabilidad de un error tipo II. La figura 10.5 presenta la curva característica de un contraste del tipo bilateral (H_0: $\vartheta = \vartheta_0$; H_1: $\vartheta \neq \vartheta_0$): cuanto menor sea α, mayor será $\beta(\vartheta)$ y al revés. La única forma de disminuir la probabilidad de ambos errores simultáneamente es aumentar el tamaño muestral.

La curva característica contiene la información más relevante del contraste, ya que determina la probabilidad de aceptar H_0 para cada valor del parámetro (es decir, cuando es cierta y cuando es falsa). En su lugar se usa también la *curva de potencia,* que indica la probabilidad complementaria de rechazar H_0 para cada valor del parámetro:

$$\text{Potencia } (\vartheta) = P(\text{rechazar } H_0 | \vartheta)$$

Figura 10.5 Curva característica del contraste H_0: $\vartheta = \vartheta_0$; H_1: $\vartheta \neq \vartheta_0$, variables normales



Dados dos contrastes definidos por dos medidas de discrepancia distintas pero con el mismo nivel de significación, escogeremos el que tenga menores probabilidades de error tipo II para cada valor del parámetro, lo que se resume diciendo que escogeremos el más potente.

Ejemplo 10.1

Se trata de contrastar con una muestra de $n = 16$ datos que la media de una población normal es $\mu = 5$ con $\sigma = 2$. La hipótesis nula es:

$$H_0 : \mu = 5$$

y supondremos que la alternativa es:

$$H_1 : \mu > 5$$

Se trata de un contraste unilateral, y tomemos como medida de discrepancia:

$$d = \frac{\bar{x} - 5}{2/\sqrt{16}} = 2(\bar{x} - 5)$$

Si H_0 es cierta, d tiene una distribución $N(0, 1)$. Tomemos $\alpha = 0{,}05$; entonces el valor d_C tal que $P(d \geq d_C | H_0) = 0{,}05$ es 1,65, y las regiones, con este nivel de significación, serán:

$$\text{Aceptar } H_0\text{: si } d \leq 1{,}65 \text{ que implica } \bar{x} \leq 5 + 1{,}66 \frac{2}{\sqrt{16}} = 5{,}825$$

$$\text{Rechazar } H_0\text{: si } d > 1{,}65 \quad ; \quad \bar{x} > 5{,}825$$

Por ejemplo, si observamos $\bar{x} = 6$ rechazaremos H_0 con una probabilidad de cometer un error tipo I de 0,05. El nivel crítico para $\bar{x} = 6$ será la probabilidad de obtener un valor mayor o igual que 6 cuando H_0 es cierta y tomamos una muestra de $n = 16$. Entonces $\bar{x} \sim N(5; 0{,}5)$, que equivale a decir que $d \sim N(0, 1)$. Como:

$$P(\bar{x} > 6 \mid \bar{x} \sim N[5, 0{,}5]) = P(d > 2 \mid N[0, 1]) = 0{,}023$$

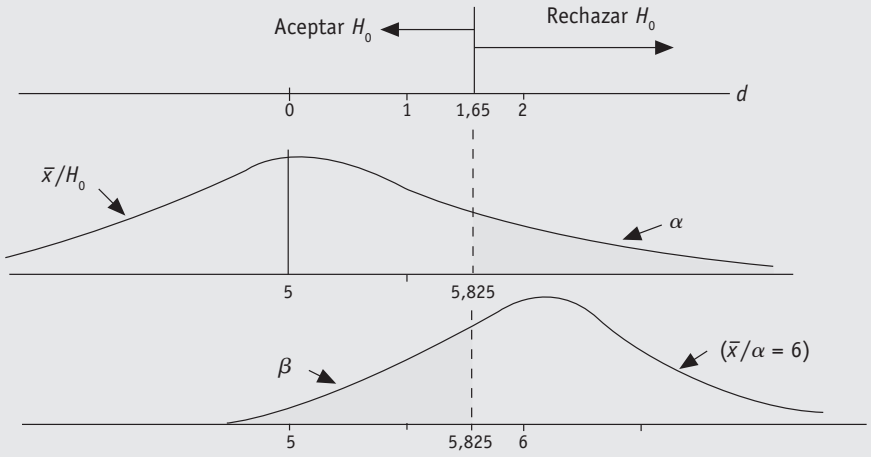
éste sería el nivel crítico del test.

Calculemos la probabilidad de un error tipo II con este contraste cuando $\mu = 6$. Entonces:

$$\beta(6) = P(d \leq 1{,}65 \mid \mu = 6) = P(\bar{x} \leq 5{,}825 \mid \mu = 6) =$$

$$= P\left(\frac{\bar{x} - 6}{0{,}5} \leq \frac{5{,}825 - 6}{0{,}5}\right) = P(z \leq -0{,}35) = 0{,}363$$

Figura 10.6 Probabilidad de cada tipo de error al realizar un contraste



La figura 10.6 ilustra esta situación: si H_0 es cierta, $\bar{x} \sim N(5; 0{,}5)$, mientras que si $\mu = 6$, $\bar{x} \sim N(6; 0{,}5)$. El área encerrada por esta segunda distribución en la región de aceptación ($\bar{x} \leq 5{,}825$) es la probabilidad de que siendo $\mu = 6$ aceptemos H_0, es decir, la probabilidad de error tipo II.

La figura muestra cómo al disminuir α aumenta β y cómo podemos reducir ambos errores simultáneamente aumentando n, lo que reduce la varianza de la distribución de d.

En general, para calcular la potencia del contraste, escribiremos:

$$\text{Pot}(\vartheta) = P(\bar{x} > 5{,}825/\bar{x} \sim N[\vartheta; 0{,}5])$$

convirtiendo $\bar{x}$ en una variable $N(0, 1)$, llamando

$$z = \frac{\bar{x} - \vartheta}{0{,}5} :$$

$$\text{Pot}(\vartheta) = P\left(z > \frac{5{,}825 - \vartheta}{0{,}5} / z \sim N[0, 1]\right)$$

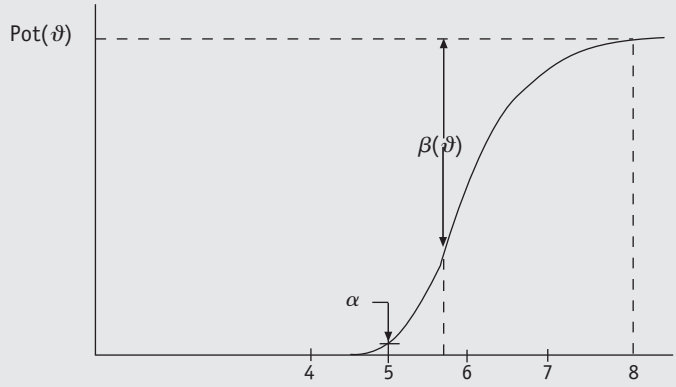
Por tanto, llamando Φ a la función de distribución de una $N(0, 1)$:

$$\text{Pot}(\vartheta) = 1 - \Phi(2[5{,}825 - \vartheta])$$

y se encuentra dibujada en la figura 10.7 a partir de la siguiente tabla:

ϑ	4	5	6	7	8
Pot(ϑ)	0,001	0,05	0,64	0,99	0,999

Figura 10.7 Curva de potencia



Ejercicios 10.1

10.1. Indique cuáles de las siguientes hipótesis son simples y cuáles compuestas: (a) $\vartheta = 20$; (b) $\vartheta \in (20, 22)$; (c) $\vartheta > 20$; (d) $\vartheta \leq 30$.

10.2. Si la distribución de una discrepancia d es una χ^2 con un grado de libertad, construir la región de rechazo correspondiente a un nivel de significación de 0,01.

10.3. La distribución de cierta medida de discrepancia es una $\chi^2(1)$. El valor observado en la muestra para ella es 5. Calcular aproximadamente el valor p crítico correspondiente.

10.4. Sugiera una medida de discrepancia para contrastar que la media de una distribución de Poisson es 5 y construya un test aproximado con $\alpha = 0{,}05$ para una muestra de tamaño 200.

10.5. Sugiera un procedimiento para contrastar que el parámetro de una distribución binomial es 0,08 con una muestra de tamaño 100 y $\alpha = 0{,}05$.

10.6. Al medir la vida media de 100 componentes se obtiene 250 horas. En la hipótesis de que la duración de vida es exponencial, contrastar la hipótesis de que la vida media de la población de componentes es 300 horas. ¿Cuál es el valor crítico p del contraste?

10.4 Contrastes para una población

10.4.1 Contraste para una proporción

Supongamos que se desea contrastar la hipótesis de que la proporción de elementos con un atributo en una población es p_0. Supondremos que la alternativa es $p \neq p_0$. Entonces, si H_0 es cierta, podemos predecir que en una muestra aleatoria de tamaño n la probabilidad de encontrar una proporción $\hat{p} = r/n$ de elementos con estas características es:

$$P\left(\hat{p} = \frac{r}{n}\right) = \binom{n}{r} p_0^r (1 - p_0)^{n-r}$$

Podemos tomar como medida de discrepancia $|p_0 - \hat{p}|$ o, lo que es lo mismo $|np_0 - r|$ siendo n el tamaño muestral y r el número de elementos con dicha característica. Para tamaños muestrales pequeños la zona de aceptación y rechazo se determinan, fijado α, por la distribución binomial (véase

ejemplo 10.2). Para tamaños muestrales grandes, utilizaremos que, si H_0 es cierta:

$$\hat{p} \sim N\left(p_0;\ \sqrt{\frac{p_0 q_0}{n}}\right)$$

en consecuencia, la región de aceptación vendrá dada por:

$$|\hat{p} - p_0| \leq z_{\alpha/2}\ \sqrt{\frac{p_0 q_0}{n}}$$

donde $z_{\alpha/2}$ es el valor correspondiente a la normal (0, 1).

Ejemplo 10.2

R. A. Fisher comenzó su famoso libro *Diseño de Experimentos* con el siguiente ejemplo: una dama afirma que el sabor de una taza de té con leche es distinto cuando se vierte antes la leche que el té. Para contrastar esta afirmación se preparan 10 tazas de té, en cinco de las cuales se vierte antes la leche y en las cinco restantes antes el té. A continuación la dama prueba, en orden aleatorio, las diez tazas —sin saber el método seguido— y acierta ocho de las diez veces. ¿Es este hecho una evidencia significativa a favor de la hipótesis?

Si el orden al mezclar los ingredientes no afecta al sabor, la probabilidad de acertar con una taza cualquiera es 0,5. El contraste será:

$$H_0: p = 0{,}5$$

$$H_1: p > 0{,}5$$

Si H_0 es cierta, la probabilidad de obtener entre 0 y 7 aciertos es (tabla 2 del apéndice):

$$P(0 \leq r \leq 7) = \sum_{0}^{7} \binom{10}{i} 0{,}5^{10} = 0{,}9452$$

Por tanto:

$$P(r > 7) = 1 - 0{,}9453 = 0{,}0547$$

En consecuencia la probabilidad de obtener más de 7 aciertos es sólo 5,5%. Por lo tanto, hay fuerte evidencia de que la dama es capaz de apreciar las diferencias de sabor.

Ejemplo 10.3

La proporción de gente que votó a un partido en unas elecciones es el 25%. Se toma hoy una muestra de $n = 500$ electores y se obtiene el 22% de votantes. ¿Hay evidencia de un cambio en el número de votos?

$$\text{Si } H_0\text{: } p = 0{,}25 \text{ es cierta, entonces } \hat{p} \sim N\left(0{,}25; \sqrt{\frac{0{,}25 \cdot 0{,}75}{500}}\right)$$

en muestras grandes, y con el 95% de probabilidad:

$$|\hat{p} - 0{,}25| < 1{,}96 \sqrt{\frac{0{,}25 \cdot 0{,}75}{500}} = 0{,}038$$

la diferencia observada, 0,03, es mayor que la esperada al 95% y corresponde a un valor de la normal (0, 1) de:

$$z_{\alpha/2} = \frac{0{,}03}{0{,}0194} = 1{,}55$$

que en las tablas de la normal proporciona un nivel crítico de $p = 2(0{,}06) = 0{,}12$. Por tanto, hay un 12% de probabilidad de encontrar discrepancias iguales o superiores a la observada. Normalmente en estos casos concluimos que no hay evidencia suficiente para suponer un cambio en los votantes.

10.4.2 Contraste de la media

Poblaciones normales

Supongamos que queremos contrastar que la media de una variable aleatoria, con distribución normal y parámetros desconocidos, es μ_0. La hipótesis nula será:

$$H_0 : \mu = \mu_0$$

frente a una hipótesis alternativa:

$$H_1 : \mu \neq \mu_0$$

Si la hipótesis nula es cierta, la media muestral $\bar{x}$ proviene de una distribución normal con media μ_0 y varianza σ desconocida. Entonces el estadístico:

$$\boxed{d = \frac{\bar{x} - \mu_0}{\hat{s}/\sqrt{n}}}$$

donde $\hat{s}$ es la desviación típica muestral corregida, tendrá una distribución t de Student con $n - 1$ grados de libertad. Por lo tanto, una región de aceptación para $\bar{x}$, al nivel de significación α, será:

$$|\bar{x} - \mu_0| \leq t_{\alpha/2} \frac{\hat{s}}{\sqrt{n}}$$

representando por $t_{\alpha/2}$ el valor de la distribución t tal que el intervalo $(-t_{\alpha/2}; t_{\alpha/2})$ contiene probabilidad $1 - \alpha$. La figura 10.8 presenta las regiones de aceptación y rechazo para este contraste. La regla de decisión será: si

$$\bar{x} \in \left(\mu_0 \pm t_{\alpha/2} \frac{\hat{s}}{\sqrt{n}} \right)$$

aceptamos H_0 con nivel de significación α, mientras que si

$$\bar{x} \notin \left(\mu_0 \pm t_{\alpha/2} \frac{\hat{s}}{\sqrt{n}} \right)$$

aceptaremos H_1.

La región de aceptación se ha tomado centrada porque la hipótesis alternativa incluye valores mayores o menores que μ_0. Siempre conviene obtener el nivel crítico del test dado por:

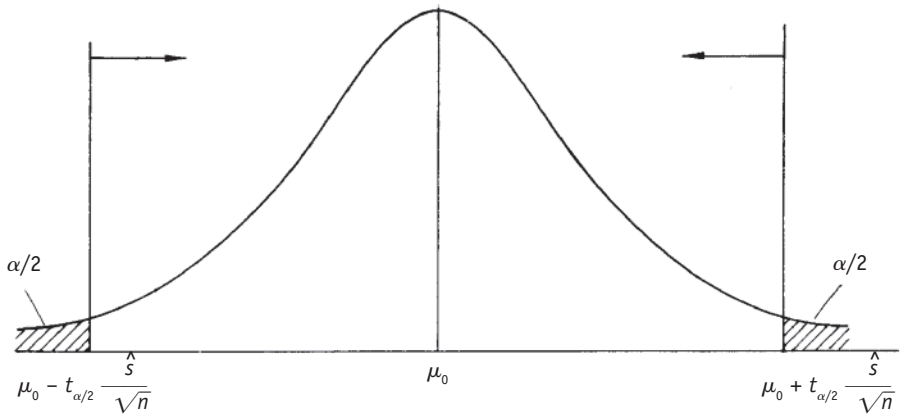
$$p = P\left(|t| > \left| \frac{\bar{x} - \mu_0}{\hat{s}/\sqrt{n}} \right| \right)$$

donde t es una variable de Student con $n - 1$ grados de libertad.

A veces se conoce que no son posibles valores de μ menores de μ_0; entonces la hipótesis alternativa se establece como:

$$H_1 : \mu > \mu_0$$

Figura 10.8 Contraste de medias



la región de rechazo sería entonces:

$$\bar{x} > \mu_0 + t_\alpha \frac{\hat{s}}{\sqrt{n}}$$

y rechazaríamos H_0 sólo para valores altos de $\bar{x}$, como sería de esperar. El nivel crítico del test será:

$$p = P\left(t > \frac{\bar{x} - \mu_0}{\hat{s}/\sqrt{n}}\right)$$

Caso general

Cuando la población es desconocida pero la muestra es grande, ($n > 30$), utilizaremos que la distribución de $\bar{x}$ es asintóticamente normal (μ, $\hat{s}/\sqrt{n}$) y tomaremos la región de aceptación como:

$$|\bar{x} - \mu_0| \leq t_{\alpha/2} \frac{\hat{s}}{\sqrt{n}}$$

10.4.3 Contraste de varianzas, poblaciones normales

Para realizar el contraste:

$$H_0 : \sigma^2 = \sigma_0^2$$
$$H_1 : \sigma^2 \neq \sigma_0^2$$

utilizaremos que $(n-1)\hat{s}^2/\sigma_0^2$ es χ^2_{n-1} y determinaremos dos valores $\chi^2_{1-\alpha/2}$ y $\chi^2_{\alpha/2}$ tales que, si H_0 es cierta, cubran el $1-\alpha$ de la distribución. La región de aceptación será:

$$\chi^2_{1-\alpha/2} \leq \frac{(n-1)\hat{s}^2}{\sigma_0^2} \leq \chi^2_{\alpha/2}$$

Si H_1 fuese $\sigma^2 > \sigma_0^2$, la región de aceptación sería del tipo:

$$\frac{(n-1)\hat{s}^2}{\sigma_0^2} \leq \chi^2_{\alpha}$$

Ejemplo 10.4

Se espera que la resistencia en kg/cm² de cierto material suministrado por un proveedor se distribuya normalmente, con media 220 y desviación típica 7,75. Se toma una muestra de 9 elementos, obteniendo:

203, 229, 215, 220, 223, 233, 208, 228, 209

Se pide:

1) Contrastar la hipótesis de que esta muestra proviene de una población con media 220 y σ cualquiera.
2) Contrastar la hipótesis de que la muestra proviene de una población con $\sigma = 7{,}75$ y media cualquiera.

Calculemos los parámetros de la muestra:

$$\bar{x} = \frac{203 + 229 + \ldots + 209}{9} = 218{,}67$$

$$\hat{s} = \left(\frac{1}{8} [(203 - 218{,}67)^2 + (229 - 218{,}67)^2 + \ldots \right.$$

$$\left. \ldots + (209 - 218{,}67]^2 \right)^{1/2} = 10{,}52$$

El contraste de la media es:

$$H_0 : \mu = 220$$
$$H_1 : \mu \neq 220$$

$$t = \frac{\bar{x} - 220}{\hat{s}/\sqrt{n}} = \frac{218{,}67 - 220}{10{,}52/\sqrt{9}} = -0{,}38$$

y aceptaremos H_0 a cualquier nivel de significación, ya que el valor de t obtenido es perfectamente consistente con H_0.

El contraste de la varianza es:

$$H_0 : \sigma^2 = 7{,}75^2$$

y tomaremos como hipótesis alternativa:

$$H_1 : \sigma^2 > 7{,}75^2$$

$$\frac{(n-1)\hat{s}^2}{\sigma^2} \text{ es } \chi^2_8 \Rightarrow \frac{8 \cdot 10{,}52^2}{7{,}75^2} = 14{,}74$$

Como $\chi^2_8(0{,}95) = 15{,}51$ y

$$\frac{(n-1)\hat{s}^2}{\sigma^2} = 14{,}74 < \chi^2_8(0{,}95) = 15{,}51$$

aceptaremos H_0 al nivel de significación de 0,05. Nótese que el valor obtenido en el test, 14,75, es alto y está cerca del valor límite. $\chi^2_8(0{,}90) = 13{,}4$, por lo que hubiéramos rechazado H_0 con nivel de significación 0,10. El nivel crítico del test es, aproximadamente, 0,10.

10.5 Comparación de dos poblaciones

10.5.1 Comparación de dos proporciones

Se desea contrastar la hipótesis de que la proporción de elementos con un atributo es idéntica en dos poblaciones. El contraste se establece:

$$H_0: p_1 = p_2 = p_0$$

$$H_1: p_1 \neq p_2$$

suponemos que se han tomado dos muestras independientes de tamaños n_1 y n_2 de ambas poblaciones obteniendo $\hat{p}_1 = r_1/n_1$ y $\hat{p}_2 = r_2/n_2$, como proporciones observadas. Si H_0 es cierta, la mejor estimación de p_0 es:

$$\hat{p}_0 = \frac{r_1 + r_2}{n_1 + n_2} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

Entonces, la variable $y = \hat{p}_1 - \hat{p}_2$ tendrá media cero y varianza igual a la suma de varianzas, dada por:

$$\text{Var}(\hat{p}_1 - \hat{p}_2) = \frac{p_0 q_0}{n_1} + \frac{p_0 q_0}{n_2}$$

En consecuencia, supuesto n_1 y n_2 grandes, la región de aceptación será:

$$|\hat{p}_1 - \hat{p}_2| \leq z_{\alpha/2}\sqrt{\frac{\hat{p}_0\hat{q}_0}{n_1} + \frac{\hat{p}_0\hat{q}_0}{n_2}} \qquad (10.1)$$

Ejemplo 10.5

La proporción de defectos en un lote de n_1 = 100 unidades del proveedor *A* es 0,04, mientras que en un lote de n_2 = 150 unidades de *B* han aparecido 0,07. ¿Hay evidencia suficiente de diferencias entre los proveedores?

$$\hat{p} = \frac{(100)0{,}04 + (150)0{,}07}{250} = 0{,}058$$

$$|\hat{p}_1 - \hat{p}_2| \leq 1{,}96\sqrt{0{,}058 \cdot 0{,}942\left(\frac{1}{100} + \frac{1}{150}\right)} = 0{,}059$$

La diferencia encontrada, 0,03, es bastante menor de la límite al 95%, con lo que no existe evidencia para suponer diferencias. La desviación observada es (0,03)/(0,0301) = 0,996, que corresponde a un nivel crítico de 0,32. Hay un 32% de probabilidad de observar discrepancias mayores que las observadas como consecuencia del azar cuando las poblaciones son idénticas.

10.5.2 Comparación de medias, varianzas iguales, muestras independientes

Dadas dos poblaciones con la misma distribución y variabilidad, pero que pueden diferir en la media, se desea contrastar la hipótesis de igualdad de medias:

$$H_0: \mu_1 = \mu_2$$

y supondremos que el contraste es bilateral,

$$H_1: \mu_1 \neq \mu_2$$

y que disponemos de dos muestras independientes de tamaños n_1 y n_2 de cada población con medias $\bar{x}_1$, $\bar{x}_2$ y desviaciones típicas $\hat{s}_1$ y $\hat{s}_2$.

Poblaciones cualesquiera

Si el tamaño muestral es grande y H_0 es cierta, el estadístico:

$$\frac{\bar{x}_1 - \bar{x}_2}{\hat{s}_T \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{10.2}$$

sigue aproximadamente una distribución $N(0, 1)$. Por tanto, la región de aceptación para un contraste bilateral será:

$$|\bar{x}_1 - \bar{x}_2| \leq z_{\alpha/2}\, \hat{s}_T \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

donde $\hat{s}_T$ es la estimación de la variabilidad común que utiliza toda la información disponible y pondera las dos estimaciones independientes $\hat{s}_1$ y $\hat{s}_2$ proporcionalmente a su precisión:

$$\hat{s}_T = \sqrt{\frac{n_1 - 1}{n_1 + n_2 - 2}\, \hat{s}_1^2 + \frac{n_2 - 1}{n_1 + n_2 - 2}\, \hat{s}_2^2} \tag{10.3}$$

Como vimos en la sección 8.6, $\hat{s}_T^2$ es la estimación centrada de σ^2 de varianza mínima.

Poblaciones normales

Si las poblaciones base son normales y H_0 es cierta, el estadístico (10.2) sigue una distribución t de Student con $n_1 + n_2 - 2$ grados de libertad, como vimos en la sección 8.6. En consecuencia, la zona de aceptación será:

$$|\bar{x}_1 - \bar{x}_2| \leq t_{\alpha/2}\, \hat{s}_T \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \tag{10.4}$$

donde $\hat{s}_T$ viene dada por (10.3). Conviene siempre calcular el nivel crítico, $P(|t| > |\hat{d}|)$, donde $\hat{d}$ se calcula con 10.2 y t es la distribución de Student con $n_1 + n_2 - 2$ grados de libertad.

10.5.3 Comparación de medias, muestras dependientes apareadas

Supongamos que queremos comparar dos marcas de neumáticos. Una posibilidad es poner durante k kilómetros la marca A en n_1 vehículos, la B en n_2, medir los desgastes medios ($\bar{x}_1$, $\bar{x}_2$) y aplicar el contraste (10.4). Como el contraste consiste en comparar las diferencias $|\bar{x}_1 - \bar{x}_2|$ con su desviación típica, cuando la variabilidad de la población sea grande, a no ser que las diferencias entre neumáticos sean enormes, o las muestras muy grandes, no las detectaremos.

El problema es que las diferencias entre los desgastes de los neumáticos dependerán de muchos factores que no controlamos y que pueden influir tanto o más que su calidad: tipo de conducción, superficie, conductor, etc. Si no controlamos estos factores, la variabilidad experimental —que estimamos por $\hat{s}_T$— será tan grande que nos impedirá observar posibles diferencias.

Una solución es disponer en cada vehículo dos neumáticos A y dos B y medir las diferencias de desgaste en el mismo vehículo. Al ser la variabilidad de estas diferencias mucho menor, tendremos, en general, un mejor contraste.

La clave del procedimiento es disponer de medidas por pares tomadas en condiciones muy semejantes, de manera que a priori las dos unidades experimentales (ruedas en el ejemplo) que comparamos sean lo más iguales posibles. De esta manera la variabilidad de las diferencias entre dos medidas será pequeña, y podemos identificar más fácilmente cambios.

Para justificar esta afirmación, supongamos que se han elegido $2n$ unidades homogéneas por pares (ruedas del mismo coche, personas de iguales

características, objetos de iguales propiedades). Sean x_{1i}, x_{2i} los valores, en el par de objetos *i*, de ambas variables. Llamando:

$$y_i = x_{1i} - x_{2i} \qquad i = 1, ..., n$$

a las diferencias en el par *i*, tendremos que:

$$E(y_i) = \mu_1 - \mu_2$$

y si no hay diferencias entre las medias, la esperanza de la diferencia es cero. Además:

$$\text{Var}(y_i) = \sigma_1^2 + \sigma_2^2 - 2\rho\ \sigma_1\sigma_2$$

donde ρ es el coeficiente de correlación entre las dos variables x_1, x_2. Suponiendo igualdad de varianzas:

$$\text{Var}(y_i) = 2\sigma^2\ (1 - \rho)$$

Si las dos medidas que comparamos (por ejemplo, los desgastes de los neumáticos en un mismo coche) son análogas, ρ será positivo y grande (próximo a uno) y la variabilidad de las desviaciones así calculadas será mucho menor que con muestras independientes.

Para realizar el contraste, estimaremos $\text{Var}(y_i)$ mediante:

$$\hat{s}_y^2 = \frac{\Sigma(y_i - \bar{y})^2}{n - 1}$$

y el contraste bilateral:

$$H_0\colon \mu_1 - \mu_2 = \mu_y = 0$$

$$H_1\colon \mu_1 \neq \mu_2\ ;\ \mu_y \neq 0$$

se efectuará de la forma habitual. La región de aceptación de H_0 será:

$$|\bar{y}| \leq t_{\alpha/2}\ \frac{\hat{s}_y}{\sqrt{n}}$$

donde la *t* tiene $n - 1$ grados de libertad.

Ejemplo 10.6

Para comparar la velocidad de dos ordenadores *A* y *B* se mide el tiempo que invierten en realizar operaciones de una cierta clase definida. Se toma una muestra de cinco operaciones de esta clase y cada operación fue realizada por ambos, obteniendo los resultados siguientes en milisegundos: *A* = (110, 125, 141, 113, 182); *B* = (102, 120, 135, 114, 175). Analizar si hay diferencias (a) teniendo en cuenta que los datos están apareados; (b) considerando muestras independientes.

Las diferencias *A*-*B* son (8, 5, 6, –1, 7), lo que supone $y = 5$; $\hat{s}_y = 3{,}53$. Entonces:

$$t = \frac{5}{3{,}53/\sqrt{5}} = 3{,}164$$

El valor *p* crítico correspondiente a 3,164 en una *t* con 4 grados de libertad es menor que 0,05 y consideramos que hay diferencias significativas entre ambos ordenadores, y que *B* es más rápido para realizar esta clase de operaciones.

Si hubiéramos supuesto muestras independientes, tendríamos:

$$\bar{x}_1 = 134{,}2 \qquad \hat{s}_1 = 29{,}37 \qquad ; \qquad \bar{x}_2 = 129{,}2 \qquad \hat{s}_2 = 27{,}3$$

Entonces:

$$s_T = \sqrt{\frac{29{,}37^2 + 27{,}3^2}{2}} = 28{,}36$$

y el contraste:

$$t_e = \frac{5}{28{,}36\sqrt{\frac{1}{5} + \frac{1}{5}}} = \frac{5}{\frac{28{,}36\sqrt{2}}{\sqrt{5}}} = 0{,}278$$

llevaría a concluir que no hay ninguna evidencia de diferencias.

10.5.4 Comparación de varianzas

Para contrastar que dos poblaciones normales tienen la misma varianza, plantearemos las hipótesis:

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_1: \sigma_1^2 \neq \sigma_2^2$$

Para construir el test, observemos que si tenemos dos muestras independientes con varianzas corregidas muestrales $\hat{s}_1^2$ y $\hat{s}_2^2$, el cociente:

$$\frac{\hat{s}_1^2}{\sigma_1^2} : \frac{\hat{s}_2^2}{\sigma_2^2} = \frac{\hat{s}_1^2}{\hat{s}_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2}$$

compara dos distribuciones χ^2 partidas por sus grados de libertad, ya que:

$$\frac{(n_1 - 1)\hat{s}_1^2}{\sigma_1^2} \text{ es } \chi^2_{n1-1} \; ; \; \frac{(n_2 - 1)\hat{s}_2^2}{\sigma_2^2} \text{ es } \chi^2_{n2-1}$$

Por tanto, el cociente anterior se distribuirá como una *F*. En la hipótesis de que $\sigma_1^2 = \sigma_2^2$, entonces:

$$d = \frac{\hat{s}_1^2}{\hat{s}_2^2} = F_{(n1-1;\, n2-1)}$$

será una *F* de Fisher con $n_1 - 1$ y $n_2 - 1$ grados de libertad. Para definir la región de aceptación buscaremos dos valores F_a y F_b tales que:

$$P(F_a \leq F \leq F_b) = 1 - \alpha$$

y el intervalo (F_a, F_b) será una región de aceptación de nivel de significación α. Es frecuente establecer este test de la forma:

$$H_0: \sigma_1^2 \leq \sigma_2^2$$
$$H_1: \sigma_1^2 > \sigma_2^2$$

El estadístico resultante será igual que en el caso anterior, pero ahora se definirá la región de aceptación buscando una F_c tal que:

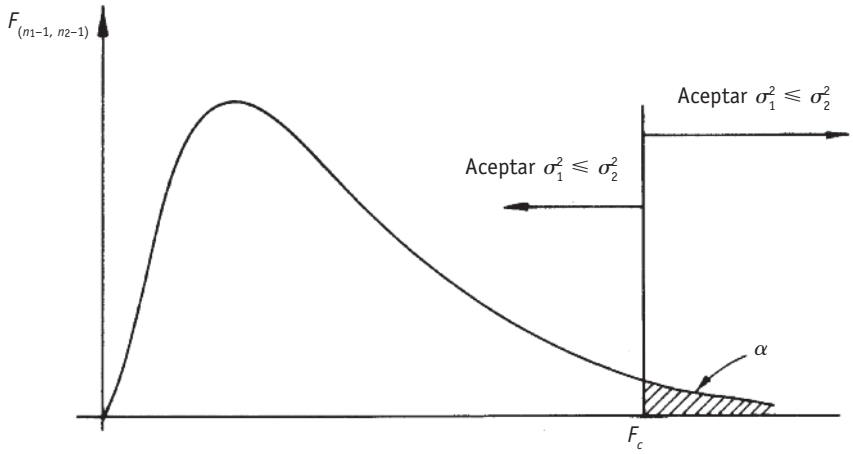
$$P(F \leq F_c) = 1 - \alpha$$

Por tanto, en este caso calcularemos el cociente:

$$F = \frac{\hat{s}_1^2}{\hat{s}_2^2}$$

y rechazaremos la hipótesis de igualdad cuando este cociente sea mayor que F_c (véase la figura 10.9).

Figura 10.9 Contraste de igualdad de varianzas, poblaciones normales



10.5.5 Comparación de medias, muestras independientes, varianzas distintas

Si las varianzas de las poblaciones son distintas, podemos utilizar el estadístico que estudiamos en el capítulo 8 para construir el intervalo en estos casos. Un enfoque alternativo es partir de que si $\sigma_2 = k\sigma_1$, y $= \bar{x}_1 - \bar{x}_2$, entonces:

$$y \sim N(\mu_1 - \mu_2;\ \sigma_1\sqrt{(1/n_1 + k^2/n_2)}$$

y llamando:

$$\hat{s}^2(k) = \frac{(n_1 - 1)\hat{s}_1^2 + (n_2 - 1)\hat{s}_2^2/k^2}{n_1 + n_2 - 2}$$

tenemos que:

$$t(k) = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\hat{s}(k)\sqrt{\dfrac{1}{n_1} + \dfrac{k^2}{n_2}}}$$

es una distribución *t* de Student con $n_1 + n_2 - 2$ grados de libertad.

Para realizar el contraste podemos suponer un valor de k (2 o 3, etc.) que deduciremos del cociente $\hat{s}_2^2/\hat{s}_1^2$, calcular $t(k)$ y obtener el nivel crítico del test, p. Variando k obtenemos distintos valores de p. Podemos encontrarnos que:

1) Para todos los valores de k el valor p es muy pequeño, con lo que rechazaremos la igualdad de medias.
2) El valor de p varía mucho con k; esto indica que cualquier conclusión que tomemos dependerá mucho de la hipótesis respecto a las varianzas.

Este procedimiento tiene las ventajas de permitir un análisis exhaustivo del problema.

Ejemplo 10.7

Se desea comparar la muestra del ejemplo 10.4 con la obtenida para otro proveedor:

221, 207, 185, 203, 187, 190, 195, 204, 212

¿Puede admitirse que ambas muestras provienen de la misma población? Comencemos comparando las varianzas. Para esta muestra:

$$\bar{x}_2 = \frac{221 + \ldots + 212}{9} = 200{,}4 \qquad \hat{s}_2 = 12{,}13$$

mientras que en el ejemplo 10.3 obtuvimos que:

$$\bar{x}_1 = 218{,}67 \qquad \hat{s}_1 = 10{,}52$$

El test será:

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2$$
$$F_{8,8} = \frac{\hat{s}_1^2}{\hat{s}_2^2} = \left(\frac{10{,}52}{12{,}13}\right)^2 = 0{,}75$$

para utilizar las tablas de una sola cola de la F, podemos colocar en el numerador siempre la varianza más grande y realizar entonces el test a una cola. Entonces:

$$F_{8,8} = \left(\frac{12{,}13}{10{,}52} \right)^2 = 1{,}33$$

y el valor crítico de F con $\alpha = 0{,}05$ es 3,44, por lo que aceptaremos la igualdad de varianzas.

El contraste de comparación de medias será:

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

la estimación de la varianza común es:

$$\hat{s}_T^2 = \frac{1}{2} (12{,}13^2 + 10{,}52^2) = 128{,}9$$
$$\hat{s}_T = 11{,}35$$

y el estadístico t será:

$$t_{16} = \frac{200{,}4 - 218{,}67}{11{,}35 \sqrt{2/9}} = -3{,}41$$

El valor crítico de t con $\alpha = 0{,}05$ es 2,12, por lo que rechazaremos la igualdad de medias con dicho valor de nivel de significación. El nivel crítico del test es aproximadamente 0,005, ya que $P(t_{16} > 3{,}25) = 0{,}0025$. Por tanto, existe una fuente evidente para rechazar H_0.

Ejemplo 10.8

Estudiar la sensibilidad de las conclusiones del ejemplo 10.7 a la hipótesis de igualdad de varianzas.

Si las varianzas fuesen σ_1^2 y $k^2\sigma_1^2$, las medias iguales, el estadístico

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{s(k) \sqrt{n_1^{-1} + k^2 n_2^{-1}}}$$

sigue una distribución t con $n_1 + n_2 - 2$ grados de libertad. Utilizando las fórmulas de la sección 10.5.5 se obtiene la tabla:

k	1	2	3	4	5	6
$s(k)$	11,35	8,59	7,97	7,74	7,63	7,58
$t(k)$	3,41	2,85	2,17	1,72	1,40	1,19

El intervalo del 99% para el cociente de varianzas proporciona en este caso un intervalo para k^2; el percentil 0,99 de una F con 8 y 8 grados de libertad que es 6,03. Por tanto:

$$\text{al } 99\% \qquad k^2 \leq 6{,}03 \Rightarrow k \leq 2{,}43$$

El valor de t obtenido para $k = 2$ es 2,85 y para $k = 3$ es 2,17; en ambos casos se rechazará la hipótesis de igualdad al 95% porque ambos valores son mayores que 2,12, valor obtenido de las tablas para una t con 16 grados de libertad. Concluimos que debemos rechazar H_0.

Ejercicios 10.2

10.2.1. Se dispone de rendimientos de dos máquinas. La máquina A ha resultado con 137,5; 140,7; 106,9; 175,1; 177,3; 120,4; 77,9 y 104,2, mientras que la B con 103,3; 121,7; 98,4; 161,5; 167,8 y 67,3. Se pide someter a contraste la hipótesis de que las máquinas son iguales con $\alpha = 0{,}05$.

10.2.2. La variabilidad de un proceso en condiciones correctas es de 3 unidades. Se dispone de una muestra de tamaño quince, con los valores siguientes: 27, 17, 18, 30, 17, 22, 16, 23, 26, 20, 22, 16, 23, 21 y 17. Se pide:
a) Contrastar la hipótesis de funcionamiento correcto con $\alpha = 0{,}05$.
b) Calcular el valor crítico p del contraste.

10.2.3. Ciertas piezas de una máquina tienen una duración media de 1.800 h. Variando uno de los materiales componentes, una muestra de 10 piezas ha dado una vida media de 2.000 horas con desviación típica de 150 horas. ¿Ha producido el material un cambio significativo de la vida de las piezas?

10.2.4. Calcular el tamaño muestral n necesario para que el contraste en una población normal de $\mu = \mu_0$ frente a $\mu = \mu_0 + \delta$ ($\delta > 0$) tenga probabilidades de error tipo I y II iguales a α.

10.2.5. Para contrastar H_0: $\mu = 1$ frente a H_1: $\mu = 2$ se dispone de una única observación, x, que proviene de una distribución de Poisson. Se toma como región crítica $x \geq 4$. Calcular las probabilidades de los errores tipo I y II.

10.2.6. Una variable x tiene la siguiente distribución de probabilidad:

	x	1	2	3	4	5	6
si H_0 es cierta	p	1/6	1/6	1/6	1/6	1/6	1/6
si H_1 es cierta	p	2/15	1/6	1/5	1/5	1/6	2/15

se decide rechazar H_0 si al observar un valor de x éste resulta ser 3 o 4. Calcular las probabilidades de ambos tipos de error y la potencia del test.

10.2.7. Utilice los datos del ejercicio 2.15 del capítulo 2 para:
a) Contrastar la hipótesis de que la variabilidad es la misma en los experimentos de Michelson y Newcomb.
b) Sabiendo que la velocidad de la luz en el vacío es 299.792,5 km/s y que la velocidad en el aire debe ser igual o menor que en el vacío, contrastar la hipótesis de que los experimentos no tienen error sistemático (es decir, la medida obtenida es igual al verdadero valor más un error aleatorio de media cero).

10.2.8. Un proceso industrial fabrica piezas con longitudes que se distribuyen normalmente (μ = 190 mm; σ = 10 mm). Se toma una muestra de tamaño cinco, obteniendo las longitudes: 187, 212, 195, 208, 192. Se pide:
a) Contrastar que estos cinco datos provienen de una población con media 190.
b) Contrastar que la varianza de la población de la cual provienen es 100.
c) Supuesto que la varianza es 100, construir la curva de potencia del contraste de la media con cinco datos y α = 0,05.

10.2.9. Una muestra de 10 piezas de acero del proveedor *A* ha dado una resistencia media a la tracción de 54.000 unidades con $\hat{s}$ = 2.100, mientras que otra muestra de 12 piezas del proveedor *B* ha resuelto en una media de 49.000 unidades y $\hat{s}$ = 1.900. Las piezas *B* son más baratas que las *A* y estas últimas sólo serían rentables si tuviesen una resistencia media de al menos 2.000 unidades mayor que *B* sin tener mayor variabilidad. En caso contrario sería mejor comprar a *B*. ¿Qué decisión se tomaría?

10.2.10. Se anuncia que el tiempo en recorrer un trayecto es por término medio de 15 horas con σ = 2 horas. Se realiza el trayecto 25 veces, obteniendo un tiempo medio de 13,8 h. Se pide:
a) Contrastar con α = 0,01 H_0: μ = 15 horas frente a μ < 15.
b) Calcular la potencia del test para α = 14 horas.
c) Dibujar la función de potencia del test.

10.2.11. Se han hecho cuatro determinaciones químicas en dos laboratorios *A* y *B* con los resultados: *A*: 26, 24, 28, 27; *B:* 28, 31, 23, 29. Suponiendo que las varianzas son iguales, contrastar con $\alpha = 0{,}05$ que los laboratorios no son significativamente distintos.

10.2.12. Un partido político afirma que el 55% de los electores están de acuerdo con él en cierto problema. Se toma una muestra de 1.000 electores y se obtiene una proporción a favor del 51%. ¿Puede el partido sostener su afirmación?

10.6 Interpretación de un contraste de hipótesis

10.6.1 Intervalos y contrastes

El cuadro 10.1 ilustra la similitud entre intervalos de confianza y contraste de hipótesis aplicado al caso de la media de una población normal. Se acepta al nivel α la hipótesis $\mu = \mu_0$ cuando el intervalo de confianza $1 - \alpha$ construido para μ incluye a μ_0 y viceversa. En general:

Intervalo de confianza $(1 - \alpha)$	=	Conjunto de hipótesis aceptables a nivel α.

Cuadro 10.1 Ejemplo comparativo de intervalos de confianza y contrastes de hipótesis

1.º) $t = \dfrac{\bar{x} - \mu}{\hat{s}/\sqrt{n}}$ tiene distribución conocida.	1.º) Si H_0: $\mu = \mu_0$, $\dfrac{\bar{x} - \mu_0}{\hat{s}/\sqrt{n}} = t$
2.º) $-t_{\alpha/2} \leq \dfrac{\bar{x} - \mu}{\hat{s}/\sqrt{n}} \leq t_{\alpha/2}$ en tablas	2.º) $-t_{\alpha/2} \leq \dfrac{\bar{x} - \mu_0}{\hat{s}/\sqrt{n}} \leq t_{\alpha/2}$ en tablas
3.º) $\mu \in \left(\bar{x} \pm t_{\alpha/2} \dfrac{\hat{s}}{\sqrt{n}} \right)$ Intervalo	3.º) $\bar{x} \in \left(\mu_0 \pm t_{\alpha/2} \dfrac{\hat{s}}{\sqrt{n}} \right)$ aceptación
Conclusión: μ está contenido en el intervalo de nivel $1 - \alpha$ si: $\lvert \bar{x} - \mu \rvert \leq t_{\alpha/2} \dfrac{\hat{s}}{\sqrt{n}}$	Conclusión: aceptamos H_0 con nivel de significación α si: $\lvert \bar{x} - \mu_0 \rvert \leq t_{\alpha/2} \dfrac{\hat{s}}{\sqrt{n}}$

En un contraste de hipótesis se define un intervalo de aceptación para el estimador; en la estimación por intervalos se invierte la relación anterior pasando a un intervalo aplicable al parámetro, y la variable ω utilizada para construir intervalos equivale a la discrepancia d utilizada en el test.

Dar un intervalo de confianza es más informativo que dar únicamente el resultado del test. Cuando exista la duda de qué procedimiento utilizar, el primero es el más recomendable.

10.6.2 Resultados significativos y no significativos

Ya hemos comentado que rechazar la hipótesis nula porque se obtiene un resultado significativo puede llevar a conclusiones absurdas en muestras grandes. En efecto, supongamos un contraste H_0: $\vartheta = \vartheta_0$ donde una muestra grande lleva a rechazar H_0, pero la mejor estimación de ϑ a la vista de la muestra es $\vartheta_0 + \varepsilon$ donde ε es muy pequeño. En este caso, desde un punto de vista práctico, los datos confirman que el parámetro está muy próximo a ϑ_0, que es realmente lo que tratamos de contrastar con H_0. En conclusión, si los datos llevan a rechazar H_0, conviene indicar siempre (1) cuál es la mejor estimación del parámetro a la vista de los datos, (2) si la diferencia es o no importante en función de la precisión de los instrumentos de medida utilizados y de la naturaleza del problema.

Análogamente, aceptar H_0 porque no se obtiene un resultado significativo puede ser de nuevo absurdo si a priori H_0 es poco verosímil y la muestra es pequeña, o la población heterogénea. Por ejemplo, supongamos una encuesta de 10 personas donde 4 de 5 mujeres y 2 de 5 hombres apoyan A. Aplicando el contraste (10.1) resulta $z = 1{,}29$ y podríamos concluir que no hay diferencias. Sin embargo, supongamos que la evidencia de otras encuestas ha indicado claramente diferencias entre los sexos como las apuntadas en esta pequeña muestra. En este caso es más razonable concluir que la muestra confirma las diferencias esperadas. La contradicción proviene de que existe información inicial fuerte que nos dice que la hipótesis H_0 de no diferencias es improbable. Si incluimos esta información con el enfoque bayesiano del capítulo 9 obtendremos un resultado acorde con el sentido común.

En otras situaciones la falta de efectos significativos es debida a la escasa potencia del contraste, consecuencia de un mal diseño de la recogida de información. Por ejemplo, con muestras apareadas pueden captarse efectos invisibles con muestras independientes (ejemplo 10.6).

En resumen, un contraste de hipótesis debe complementarse siempre con la estimación de los parámetros y un análisis de potencia para evaluar su capacidad de discriminación. La aplicación mecánica de esta herramienta no es recomendable.

10.7 Contrastes de la razón de verosimilitudes

10.7.1 Introducción

La teoría de contrastes de hipótesis mediante la razón de verosimilitudes fue expuesta por Neyman y Pearson y presenta las ventajas siguientes:

1) Ofrece un procedimiento para diseñar y comparar nuevos contrastes.
2) Pone de manifiesto el papel central de la función de verosimilitud en cualquier proceso de inferencia.
3) Proporciona contrastes asintóticos que pueden aplicarse en una amplia gama de situaciones, donde es difícil disponer de contrastes exactos.
4) Permite construir contrastes de hipótesis para vectores de parámetros.

10.7.2 Contraste de hipótesis simple frente alternativa simple

Supongamos que se trata de contrastar H_0: $\vartheta = \vartheta_0$ frente a H_1: $\vartheta = \vartheta_1$. Sea $\ell(\vartheta|\mathbf{X})$ la función de verosimilitud para ϑ dada la muestra. Si

$$\ell(\vartheta_1|\mathbf{X}) > \ell(\vartheta_0|\mathbf{X})$$

los datos apoyan más a H_1 que a H_0. Podríamos pues rechazar H_0 cuando el cociente

$$\lambda = \frac{\ell(\vartheta_1)}{\ell(\vartheta_0)}$$

fuese suficientemente grande. (Recordemos que las magnitudes relevantes son los cocientes y no las diferencias de verosimilitudes, que pueden combinarse arbitrariamente multiplicando por constantes.)

Un procedimiento para construir contrastes es tomar como medida de discrepancia el estadístico λ. Si la distribución de λ cuando H_0 es cierta es conocida, el contraste queda automáticamente determinado al fijar α, tomando como región crítica el conjunto:

$$\lambda > \lambda_C$$

donde λ_C se determina por:

$$P(\lambda > \lambda_C/H_0) = \alpha$$

A veces es más cómodo trabajar con las diferencias de soportes:

$$\ln \lambda = L(\vartheta_1) - L(\vartheta_0)$$

que podemos tomar también como medida de discrepancia para realizar el test.

Neyman y Pearson demostraron que, en el caso de hipótesis simple frente a alternativa simple, este procedimiento proporciona automáticamente el contraste más potente (óptimo). Si existe un estadístico suficiente, la función de verosimilitud, y por tanto λ, es función de él y el contraste de la razón de verosimilitudes conduce a comparar el estimador máximo-verosímil (que es función del suficiente, si existe) con el parámetro.

Ejemplo 10.9

Se desea contrastar que la proporción de piezas defectuosas en un proceso es 0,02 frente a la hipótesis alternativa $p = 0{,}05$. Diseñar un contraste de razón y verosimilitudes para $\alpha = 0{,}05$. Aplicación para $n = 100$.

$$H_0 : p = 0{,}02 = p_0$$
$$H_1 : p = 0{,}05 = p_1$$

Al tomar una muestra de tamaño n, la verosimilitud será $\ell(p, x) = p^{\Sigma x_i} q^{n-\Sigma x_i}$. Por tanto, llamando $\Sigma x_i = r$

$$\lambda = \frac{p_1^r q_1^{n-r}}{p_0^r q_0^{n-r}} = \left(\frac{5}{2}\right)^r \left(\frac{95}{98}\right)^{n-r}$$

y la condición $\lambda > \lambda_c$ equivale a $r > k$, que será el contraste de la razón de verosimilitudes. La constante k se determinará fijando el nivel de significación:

$$\alpha = P(r > k \,/\, p_0)$$

o utilizando la aproximación normal de la binomial para n grande:

$$\alpha = P(r > k \,/\, r \sim N[np_0 \quad ; \quad \sqrt{np_0q_0}])$$

Particularizando para $n = 100$ y $p_0 = 0{,}02$ tendremos que r es $N(2; 1{,}4)$; por tanto:

$$P\left(\frac{r-2}{1,4} > \frac{k-2}{1,4}\right) = 0,05$$

en tablas de la normal estándar

$$\frac{k-2}{1,4} = 1,65$$

$$k = 4,31$$

y el test será rechazar H_0 si $r > 4,3$, es decir, si $r \geq 5$. Por tanto, no es posible encontrar un test con exactamente $\alpha = 0,05$; el test más próximo será con $k > 5$, es decir, rechazando que $p = 0,02$ cuando hay un 5% o más de piezas defectuosas. El nivel de significación será:

$$P(r \geq 5) = P(r > 4,5) = P(z \geq 1,786) = 0,94$$

Por tanto, $\alpha \simeq 0,06$ y el test es: si $r < 5$, aceptar $p = 0,02$; si $r \geq 5$ aceptar $p = 0,05$.

10.7.3 Contrastes de hipótesis compuestas

Cuando alguna o ambas de las dos hipótesis son compuestas, el procedimiento anterior no puede aplicarse directamente, ya que la función de verosimilitud de una hipótesis compuesta queda indeterminada.

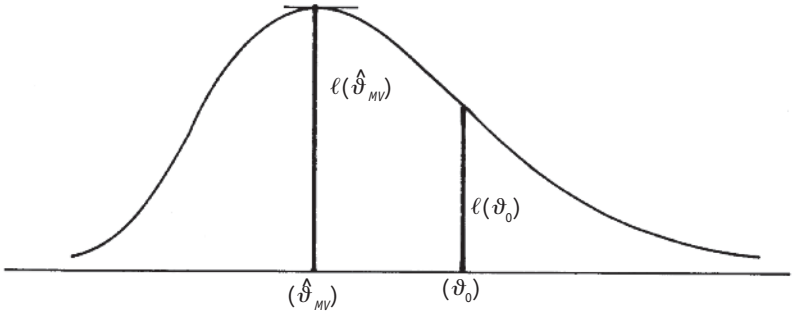
Generalizando, consideremos el caso:

$$H_0: \vartheta = \vartheta_0$$
$$H_1: \vartheta \neq \vartheta_0$$

Ahora, aunque la razón de verosimilitud para H_0, $\ell(\vartheta_0)$, está bien definida, $\ell(\vartheta_1)$ no lo está. Para representar H_1 podemos tomar el valor más favorable compatible con ella, que será el que conduzca a una verosimilitud mayor. La figura 10.10 representa esta situación.

En consecuencia, representaremos H_1 por $\vartheta = \hat{\vartheta}_{MV}$, el estimador máximo-verosímil, que hace máxima la función $\ell(\vartheta)$. El contraste será pues:

Figura 10.10 Elección de un valor representativo para H_1 en la función de verosimilitud



$$\lambda = \frac{\max_{\vartheta \neq \vartheta_0} \ell(\vartheta)}{\ell(\vartheta_0)} = \frac{\ell(\hat{\vartheta}_{MV})}{\ell(\vartheta_0)}$$

Así construido λ es siempre mayor que uno. Rechazaremos H_0 cuando (figura 10.11):

$$\lambda > \lambda_C$$

donde:

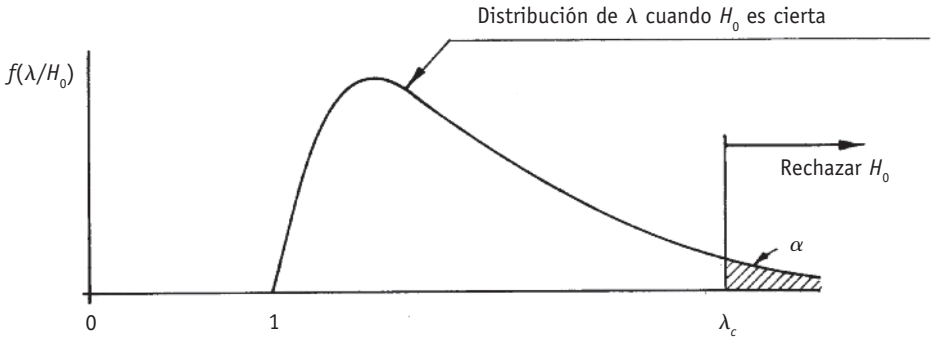
$$P(\lambda > \lambda_C / H_0) = \alpha$$

Frecuentemente la distribución $f(\lambda/H_0)$ no es directamente conocida, pero sí lo es la de $\log \lambda$ u otra función de λ, con lo que podemos determinar k. En estos casos, si g es una función biunívoca de λ el contraste se establece como:

$$g(\lambda) = g\left(\frac{\ell[\hat{\vartheta}_{MV}]}{\ell[\vartheta_0]}\right) > k$$

Aplicando este método a los problemas de contraste uniparamétricos antes estudiados, se obtienen las medidas de discrepancia introducidas previamente en términos intuitivos. La razón es simple, $g(\lambda)$, puede escribirse como $g(\hat{\vartheta}_{MV}; \vartheta_0)$, y medirá la discrepancia relativa entre $\hat{\vartheta}_{MV}$ y ϑ_0.

Figura 10.11 Definición de la región crítica o de rechazo



En el caso en que H_0 sea compuesta, si el contraste es:

$$H_0: \vartheta \in (a, b)$$
$$H_1: \vartheta \notin (a, b)$$

tomaremos como valor representativo de H_0 el máximo de la función de verosimilitud en cada zona:

$$\lambda = \frac{\max\limits_{\vartheta \notin (a,b)} \ell(\vartheta)}{\max\limits_{\vartheta \in (a,b)} \ell(\vartheta)} = \frac{\max \ell(H_1)}{\max \ell(H_0)}$$

Este último caso es importante cuando ϑ es un vector, como veremos en la sección siguiente.

Ejemplo 10.10

Construir un contraste de razón de verosimilitudes para la hipótesis:

$$H_0: \mu = \mu_0 \qquad \text{frente a} \qquad H_1: \mu \neq \mu_0$$

en poblaciones normales con σ conocido. Entonces:

$$\ell(\mu) = k \cdot e^{-n\frac{(\bar{x} - \mu)^2}{2\sigma^2}}$$

que implica:

$$\ell(\mu_0) = k \cdot e^{\frac{-n}{2\sigma^2}(\bar{x} - \mu_0)^2}$$

$$\ell(\hat{\mu}_{MV} = \bar{x}) = k$$

Por tanto:

$$\lambda = \frac{\ell(\bar{x})}{\ell(\mu_0)} = e^{\frac{n}{2\sigma^2}(\bar{x} - \mu_0)^2}$$

La distribución de λ no es inmediata, pero tomando $g = 2 \ln \lambda$, resulta:

$$g(\lambda) = 2 \ln \lambda = \frac{n}{\sigma^2}(\bar{x} - \mu_0)^2$$

que si H_0 es cierta tiene una distribución χ^2 con 1 grado de libertad. Por tanto, el contraste será rechazar H_0 cuando:

$$\left(\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right) > c$$

es decir, cuando:

$$-c' < \left(\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right) < c'$$

donde $c' = \sqrt{c}$ se determina imponiendo que la discrepancia relativa $\sqrt{n}(\bar{x} - \mu_0)/\sigma$ es una $N(0, 1)$. Éste es el test que obtuvimos anteriormente.

10.7.4 Contrastes para varios parámetros

Introducción

Es frecuente que necesitemos contrastar una hipótesis respecto a un vector de parámetros. Por ejemplo, en el contraste de la media de una población normal con varianza desconocida, existen dos parámetros desconocidos: μ y σ^2, y el contraste de un valor particular de μ se establece:

$$H_0: \mu = \mu_0 \qquad \sigma^2 > 0$$

$$H_1: \mu \neq \mu_0 \qquad \sigma^2 > 0$$

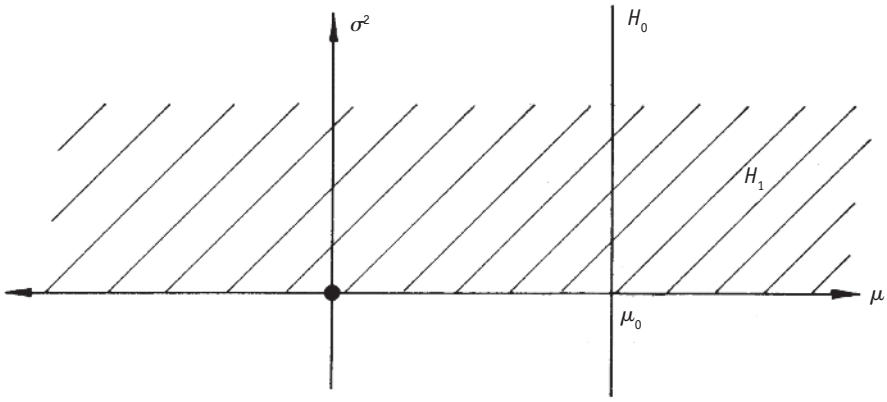
y afecta al vector de parámetros $\boldsymbol{\vartheta}' = (\mu, \sigma^2)$. Llamaremos *espacio paramétrico* de $\boldsymbol{\vartheta}$ al espacio de dimensión dos que incluye todos sus posibles valores. Este espacio está definido en el ejemplo anterior por:

$$-\infty < \mu < \infty$$

$$\sigma^2 > 0$$

La hipótesis H_0 especifica que $\boldsymbol{\vartheta}'$ pertenece a un espacio de dimensión uno: la recta $\mu = \mu_0$; mientras que H_1 no establece restricciones sobre los posibles valores de $\boldsymbol{\vartheta}$. La figura 10.12 ilustra gráficamente las dos regiones asociadas a las hipótesis.

Figura 10.12 Contraste de la media, σ^2 desconocido



Análogamente, el contraste de igualdad de medias de dos poblaciones con varianzas desconocidas, aunque iguales, afecta al vector de dimensión tres $\boldsymbol{\vartheta}' = (\mu_1, \mu_2, \sigma^2)$:

$$H_0: \mu_1 - \mu_2 = 0; \quad \sigma^2 > 0$$

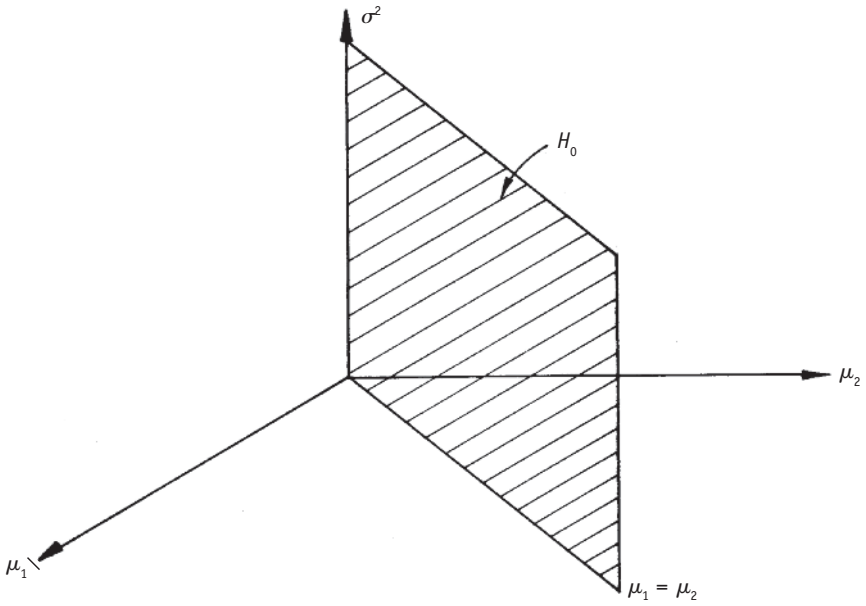
frente a:

$$H_1: \mu_1 \text{ y } \mu_2 \text{ cualesquiera}, \sigma^2 > 0.$$

H_0 restringe al vector ϑ a pertenecer a un plano; mientras que H_1 permite que ϑ sea cualquier punto del espacio paramétrico (véase la figura 10.13).

Estos ejemplos ilustran la necesidad de un enfoque general para estos problemas.

Figura 10.13 Contraste de la igualdad de medias



El contraste general de razón de verosimilitudes

Supongamos que $\boldsymbol{\vartheta}$ es un vector p-dimensional y se pretende contrastar:

$$H_0: \boldsymbol{\vartheta} \in \Omega_0$$

donde Ω_0 es un subconjunto de valores posibles del espacio p-dimensional de valores de ϑ, Ω, frente a la alternativa

$$H_1: \boldsymbol{\vartheta} \in \Omega - \Omega_0$$

En la mayoría de los casos de interés la hipótesis H_0 establece un conjunto de restricciones sobre todos, o una parte, de los componentes de $\boldsymbol{\vartheta}' = (\vartheta_1, ..., \vartheta_p)$. Por ejemplo:

$$\text{(a) } \boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0; \quad \text{(b) } \boldsymbol{\vartheta} = \mathbf{A}\boldsymbol{\beta}; \quad \text{(c) } \vartheta_1 = ... = \vartheta_r = 0; \quad \text{(d) } \mathbf{a}'\boldsymbol{\vartheta} = 0$$

La hipótesis (a) indica que el vector de parámetros es igual a un valor dado (el conjunto Ω_0 se reduce pues a un punto); la (b) establece que si **A** es una matriz $p \times h$ ($h < p$) de coeficientes conocidos y rango h, los p parámetros pueden expresarse como una función lineal de otros h parámetros $\boldsymbol{\beta}$ (el espacio Ω está definido por el hiperplano de dimensión h generado por los vectores columna de la matriz **A**); la hipótesis (c) fija los valores de r componentes (la dimensión de Ω_0 es $p - r$); finalmente, la hipótesis (d), donde **a** es un vector de constantes dado, establece una relación lineal entre los p coeficientes de $\boldsymbol{\vartheta}$ (la dimensión de Ω_0 es $p - 1$).

La hipótesis alternativa, H_1, para los cuatro casos anteriores es que $\boldsymbol{\vartheta}$ no está sujeta a estas restricciones; por tanto, siempre la dimensión del espacio $\Omega - \Omega_0$ es p.

Es posible que el valor de la función verosimilitud para H_0 no quede definido. Por ejemplo, la hipótesis (a) anterior especifica un valor único para la función, pero las restantes permiten un conjunto de valores posibles para ϑ, ya que son hipótesis compuestas.

Es razonable caracterizar H_0 por el valor máximo de la verosimilitud en el conjunto que define la hipótesis, Ω_0. Análogamente, H_1 se caracteriza por el valor máximo en $\Omega - \Omega_0$ que, en los casos de interés, será igual al máximo en todo el espacio, ya que, normalmente, la región definida por Ω_0 es un subconjunto muy pequeño del espacio paramétrico. Con estas hipótesis, el ratio de verosimilitudes asociado a las hipótesis será:

$$\lambda = \frac{\ell(\hat{\Omega})}{\ell(\hat{\Omega}_0)}$$

donde $\ell(\hat{\Omega})$ corresponde al máximo en todo el espacio y $\ell(\hat{\Omega}_0)$ al máximo de la verosimilitud restringida al espacio $\hat{\Omega}_0$. Es obvio que $\lambda \geq 1$, y la región de rechazo vendrá definida por valores grandes de λ, es decir:

$$\text{región crítica: } \lambda > \lambda_c$$

donde λ_c se determinará, como en la sección anterior, imponiendo que el nivel de significación del test sea α. Para ello, es necesario conocer la distribución de λ cuando H_0 es cierta, lo que suele ser difícil en la práctica. Sin embargo, cuando el tamaño muestral es grande, la distribución de $2\ln\lambda$ cuando H_0 es cierta, definida por:

$$2\ln\lambda = 2[L(\hat{\Omega}) - L(\hat{\Omega}_0)]$$

donde L es la función soporte, será, asintóticamente, una x^2 con un número de grados de libertad igual a la diferencia de dimensión entre los espacios

$\Omega - \Omega_0$, y Ω_0, o, lo que es lo mismo, igual al número de restricciones lineales impuesto por H_0.

En general la dimensión de $\Omega - \Omega_0$ será p, y la dimensión de Ω_0 será $p - r$, siendo r el número de restricciones lineales sobre el vector de parámetros; por tanto:

$$g = gl(2\ln\lambda) = \dim(\Omega - \Omega_0) - \dim(\Omega_0) = p - (p - r) = r$$

Interpretación

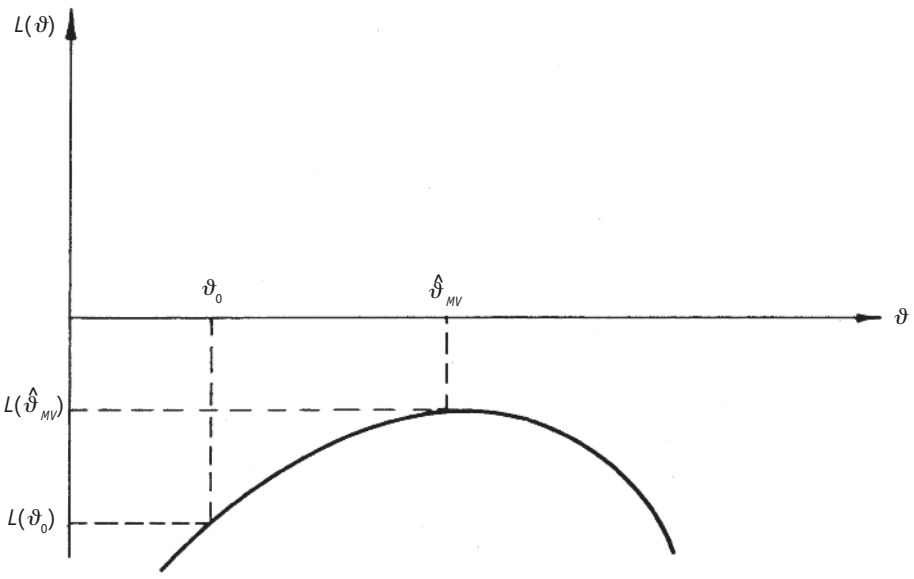
La figura 10.14 ilustra gráficamente el contraste de la razón de verosimilitudes en el caso de un parámetro.

Si $\hat{\vartheta}_{MV}$ está próximo a ϑ_0, las verosimilitudes en ambos puntos serán análogas. Desarrollando en serie $L(\vartheta)$ alrededor de $\hat{\vartheta}_{MV}$

$$L(\vartheta) \simeq \frac{-1}{2}\left[\frac{\vartheta - \hat{\vartheta}_{MV}}{\hat{\sigma}(\hat{\vartheta}_{MV})}\right]^2 + L(\hat{\vartheta}_{MV})$$

donde hemos utilizado que la primera derivada del soporte es nula en $\hat{\vartheta}_{MV}$ y que la segunda es la inversa cambiada de signo de la varianza del estimador *MV*.

Figura 10.14 La función soporte y el test de verosimilitudes



Particularizando la expresión anterior para $\vartheta = \vartheta_0$:

$$2[L(\hat{\vartheta}_{MV}) - L(\vartheta_0)] \simeq \left[\frac{\vartheta_0 - \hat{\vartheta}_{MV}}{\hat{\sigma}(\hat{\vartheta}_{MV})}\right]^2$$

y la razón de verosimilitud mide la distancia estandarizada entre ϑ_0 y $\hat{\vartheta}_{MV}$ (véase el apéndice 10B).

Ejemplo 10.11

Como ilustración del método expuesto, vamos a deducir el contraste para la media de una población normal con σ desconocida. La hipótesis nula será:

$$H_0: \mu = \mu_0, \qquad \sigma^2 > 0$$

y la alternativa:

$$H_1: \mu \neq \mu_0, \qquad \sigma^2 > 0$$

La función soporte para la muestra es:

$$L(\mu, \sigma^2) = -\frac{n}{2} \ln \sigma^2 - \frac{1}{2} \frac{\Sigma(x_i - \mu)^2}{\sigma^2}$$

y para obtener su máximo en Ω_0, sustituyendo μ_0, y derivando respecto a σ^2:

$$\frac{\partial L(\mu_0, \sigma^2)}{\partial \sigma^2} = 0 = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2}\frac{\Sigma(x_i - \mu_0)^2}{\sigma^4}$$

que conducirá a:

$$\hat{\sigma}_0^2 = \frac{1}{n}\Sigma(x_i - \mu_0)^2$$

y, por tanto:

$$L(\hat{\Omega}_0) = -\frac{n}{2} \ln \hat{\sigma}_0^2 - \frac{n}{2}$$

Análogamente, sustituyendo los estimadores máximo-verosímiles:

$$L(\hat{\Omega}) = -\frac{n}{2}\ln \hat{s}^2 - \frac{n}{2}$$

resulta que:

$$2\ln\lambda = -n\ln\hat{s}^2 + n\ln\hat{\sigma}_0^2 = n\ln\frac{\hat{\sigma}_0^2}{\hat{s}^2}$$

será una χ^2 con 1 grado de libertad. Vamos a comprobar que este contraste asintótico coincide, para n grande, con el estudiado en la sección 10.4.2. Como:

$$\frac{\hat{\sigma}_0^2}{\hat{s}^2} = \frac{\Sigma(x_i - \mu_0)^2}{n\hat{s}^2} = \frac{\Sigma(x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2}{n\hat{s}^2}$$

tendremos que:

$$\frac{\hat{\sigma}_0^2}{\hat{s}^2} = 1 + \left(\frac{\bar{x} - \mu_0}{\hat{s}}\right)^2$$

Como la varianza de $\bar{x}$ es σ^2/n, el cociente $[(\bar{x} - \mu_0)/\hat{s}]^2$, cuando H_0 es cierta, será pequeño, incluso para tamaños de muestra pequeños. Entonces, como cuando x es pequeño:

$$\ln(1 + x) \simeq x$$

concluimos que:

$$2\ln\lambda = n\ln\left[1 + \left(\frac{\bar{x} - \mu_0}{\hat{s}}\right)^2\right] = \frac{n(\bar{x} - \mu_0)^2}{\hat{s}^2}$$

o, lo que es lo mismo, el test puede hacerse con el estadístico:

$$\frac{\sqrt{n}(\bar{x} - \mu_0)}{\hat{s}}$$

que, si la hipótesis H_0 es cierta, se distribuirá, como sabemos, como una t de Student.

En conclusión, vemos que el contraste general de la razón de verosimilitudes, que es un contraste asintótico, será próximo, cuando n sea gran-

de, al contraste exacto de la t de Student. En efecto, si n es grande, como $\hat{s}^2$ es un estimador consistente de σ^2:

$$t = \frac{\sqrt{n}(\bar{x} - \mu_0)}{\hat{s}} \simeq \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} = z$$

y la variable t será aproximadamente una $N(0, 1)$.

Ejemplo 10.12

Deducir el contraste de razón de verosimilitudes en el problema:

$$H_0: \mu_1 = \mu_2 = \ldots = \mu_k;\ \sigma^2 > 0$$

$$H_1: \text{medias cualesquiera},\ \sigma^2 > 0$$

cuando tenemos k muestras de tamaños $n_1, \ldots, n_k$ de k poblaciones normales con la misma varianza. La función soporte será:

$$L(\mu_1, \mu_2, \ldots, \mu_k, \sigma^2) = -\frac{1}{2}\Sigma n_i \ln \sigma^2 - \frac{1}{2\sigma^2}\Sigma\Sigma(x_{ij} - \mu_i)^2$$

donde x_{ij} representa la observación j, ($j = 1, \ldots, n_i$), de la población i. Entonces, llamando $n = \Sigma n_i$,

$$L(\hat{\Omega}_0) = -\frac{n}{2}\ln \hat{\sigma}_0^2 - \frac{n}{2}$$

donde:

$$\hat{\sigma}_0^2 = \frac{\Sigma\Sigma(x_{ij} - \bar{x})^2}{n}$$

y

$$\bar{x} = \frac{\Sigma\Sigma x_{ij}}{n}$$

mientras que:

$$L(\hat{\Omega}) = -\frac{n}{2}\ln \hat{\sigma}_i^2 - \frac{n}{2}$$

donde:

$$\hat{\sigma}_i^2 = \frac{\sum_i \sum_j (x_{ij} - \bar{x}_i)^2}{n}$$

con:

$$\bar{x}_i = \frac{\Sigma x_{ij}}{n_i}$$

y, por tanto:

$$2 \ln \lambda = n \ln \frac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2}$$

y la variable $n \ln(\hat{\sigma}_0^2/\hat{\sigma}_1^2)$ será una χ^2; como la dimensión de H_0 es igual al número de parámetros, $k + 1$, y la de H_0 es dos al existir las $k - 1$ restricciones:

$$\mu_1 - \mu_2 = 0$$
$$\mu_2 - \mu_3 = 0$$
$$\ldots\ldots\ldots\ldots\ldots$$
$$\mu_{k-1} - \mu_k = 0$$

la dimensión de χ^2 resultante es $k - 1$.

Ejercicios 10.3

10.3.1. Contrastar $\vartheta = 1$ frente a $\vartheta > 1$ mediante la razón de verosimilitudes para $f(x) = e^{-x/\vartheta}/\vartheta$ $(x \geq 0)$. Tomar $\alpha = 0{,}05$, $n = 10$.

10.3.2. Deducir un contraste aproximado de la razón de verosimilitudes de que el coeficiente de correlación de una distribución normal bivariante es cero. Aplicación a $\alpha = 0{,}05$, $n = 20$.

10.3.3. Deducir los contrastes para poblaciones normales de la sección 4 como contrastes de la razón de verosimilitudes.

10.3.4. Deducir el contraste de la razón de verosimilitudes para $p = p_0$ frente a $p \neq p_0$ en una población binomial.

10.3.5. Establecer un contraste de la razón de verosimilitudes para $\vartheta = 2$ frente a $\vartheta > 2$ en $f(x) = 2x/\vartheta^2\ (\vartheta \leq x \leq \vartheta)$.

10.8 Resumen del capítulo

Este capítulo ha presentado la metodología estadística para contrastar hipótesis. Los contrastes principales obtenidos para una y dos poblaciones se resumen en el cuadro 10.2. Los contrastes para comparar varias poblaciones cuantitativas en el segundo tomo. La idea central es siempre la misma: comparar las predicciones generadas por la hipótesis con los datos observados y rechazar la hipótesis si la discrepancia es demasiado grande para poder ser atribuida al azar. Para evaluar la eficacia de un contraste hay que conocer las probabilidades de los dos tipos de error que podemos cometer: rechazar la hipótesis nula cuando es cierta (error tipo I) y aceptar la hipótesis nula cuando es falsa (error tipo II). Los contrastes se construyen generalmente fijando la probabilidad de un error tipo I y, siempre que sea posible, conviene calcular su potencia, que proporciona la probabilidad de rechazar la hipótesis nula para cualquier valor del parámetro.

10.9 Lecturas recomendadas

El contraste de hipótesis se trata en todos los manuales básicos de estadística que se citan en la bibliografía. El lector interesado en una introducción simple puede acudir a Wonacott y Wonnacott (2004) y Newbold *et al.* (2006). El material aquí expuesto puede ampliarse en Cox y Hinkely (1979), De Groot (1988) y Lindgren (1993). Silvey (1975) es especialmente claro y recomendable. Una referencia clásica del enfoque de Neyman-Pearson es Lehmann y Casella (2003).

Cuadro 10.2 Resumen de los contrastes principales presentados en el capítulo

Población	Parámetro	Contraste	Región de aceptación
a) Contraste para una población			
Binomial	p	H_0: $p = p_0$ H_1: $p \neq p_1$	$\lvert\hat{p} - p_0\rvert < z_{\alpha/2}\sqrt{\dfrac{p_0 q_0}{n}}$
Normal o muestras grandes	μ	H_0: $\mu = \mu_0$ H_1: $\mu \neq \mu_0$	$\lvert x - \mu_0\rvert \leq t_{\alpha/2}\dfrac{\hat{s}}{\sqrt{n}}$
Normal	σ^2	H_0: $\sigma^2 = \sigma_0^2$ H_1: $\sigma^2 > \sigma_0^2$	$\dfrac{(n-1)\hat{s}^2}{\sigma_0^2} \leq \chi^2\,(n-1;\,\alpha)$
Cualquiera	$\boldsymbol{\vartheta}$	H_0: $\boldsymbol{\vartheta} \in \Omega_0$ H_1: $\boldsymbol{\vartheta} \in \Omega - \Omega_0$	$2 \ln \lambda = 2[L(\hat{\Omega}) - L(\hat{\Omega}_0)] < \chi_g^2$
b) Contrastes para dos poblaciones			
Binomiales	p_1, p_2	H_0: $p_1 = p_2$ H_1: $p_1 \neq p_2$	$\lvert\hat{p}_1 - \hat{p}_2\rvert < z_{\alpha/2}\sqrt{\hat{p}_0\hat{q}_0\dfrac{1}{n_1} + \dfrac{1}{n_2}}$ $\hat{p}_0 = \alpha\hat{p}_1 + (1-\alpha)\hat{p}_2$ $\alpha = n_1/(n_1 + n_2)$
Normales con misma σ^2	μ_1, μ_2	H_0: $\mu_1 = \mu_2$ H_1: $\mu_1 \neq \mu_2$	$\lvert\bar{x}_1 - \bar{x}_2\rvert \leq z_{\alpha/2}\,\hat{s}_T\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$ $\hat{s}_T = \sqrt{\alpha s_1^2 + (1-\alpha)\hat{s}_2^2}$ $\alpha = (n_1 - 1)/(n_1 + n_2 - 2)$
Normales apareadas	μ_1, μ_2	H_0: $\mu_1 = \mu_2$ H_1: $\mu_1 \neq \mu_2$	$\lvert\bar{y}\rvert \leq t_{\alpha/2}\dfrac{\hat{s}_y}{\sqrt{n}}$ $y = x_1 - x_2$
Normales	σ_1^2, σ_2^2	H_0: $\sigma_1^2 = \sigma_2^2$ H_1: $\sigma_1^2 > \sigma_2^2$	$\hat{s}_1^2/\hat{s}_2^2 \leq F(n_1 - 1, n_2 - 1;\,\alpha)$
Normales distinta σ^2	μ_1, μ_2	H_0: $\mu_1 = \mu_2$ H_1: $\mu_1 \neq \mu_2$	$\lvert\bar{x}_1 - \bar{x}_2\rvert \leq t(k)s(k)\sqrt{\dfrac{1}{n_1} + \dfrac{k^2}{n_2}}$ $k = \sigma_2/\sigma_1$

Apéndice 10A: Deducción del contraste de verosimilitudes

Vamos a indicar las líneas generales de la demostración de que $2\ln\lambda$ se distribuye como una χ^2 cuando H_0 es cierta. Sea $L(\boldsymbol{\vartheta})$ la función soporte. Desarrollemos esta función en un entorno del estimador máximo-verosímil $\hat{\boldsymbol{\vartheta}}_{MV}$. Tendremos:

$$L(\boldsymbol{\vartheta}) \simeq L(\hat{\boldsymbol{\vartheta}}_{MV}) + \nabla \mathbf{L}(\hat{\boldsymbol{\vartheta}}_{MV})'(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}_{MV}) + \frac{1}{2}(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}_{MV})'\mathbf{H}(\hat{\boldsymbol{\vartheta}}_{MV})(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}_{MV})$$

donde $\nabla \mathbf{L}(\hat{\boldsymbol{\vartheta}}_{MV})$ representa el vector de primeras derivadas, y $\mathbf{H}(\hat{\boldsymbol{\vartheta}}_{MV})$ la matriz (hessiana) de segundas derivadas, ambos evaluados en el punto $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_{MV}$; como, por definición, las primeras derivadas son nulas en el máximo, y la matriz hessiana es, en ese punto, definida negativa:

$$L(\boldsymbol{\vartheta}) = L(\hat{\boldsymbol{\vartheta}}_{MV}) - \frac{1}{2}(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}_{MV})'[-\mathbf{H}(\hat{\boldsymbol{\vartheta}}_{MV})](\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}_{MV}) \qquad (10A.1)$$

donde ahora la matriz $-\mathbf{H}(\hat{\boldsymbol{\vartheta}}_{MV})$ es definida positiva.

Sea $\boldsymbol{\vartheta}_0$ el estimador obtenido maximizando la función de verosimilitud en el conjunto Ω_0. El máximo en el espacio Ω es el estimador máximo-verosímil, que corresponde al máximo sin ningún tipo de restricción. El contraste de razón de verosimilitudes, *TV*, es:

$$TV = 2 \ln \lambda = 2[L(\hat{\boldsymbol{\vartheta}}_{MV}) - L(\hat{\boldsymbol{\vartheta}}_0)]$$

Supongamos ahora que H_0 es cierta: es de esperar que $\hat{\boldsymbol{\vartheta}}_0$, el estimador obtenido imponiendo la restricción cierta de que $\boldsymbol{\vartheta}$ pertenezca a Ω_0, debe ser similar, asintóticamente, al estimador $\hat{\boldsymbol{\vartheta}}_{MV}$ obtenido sin estas restricciones. Sustituyendo la diferencia de soportes por el desarrollo (10.A.1) particularizado para $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$, tendremos:

$$TV = 2 \ln \lambda = (\hat{\boldsymbol{\vartheta}}_0 - \hat{\boldsymbol{\vartheta}}_{MV})'[-\mathbf{H}(\hat{\boldsymbol{\vartheta}}_{MV})](\hat{\boldsymbol{\vartheta}}_0 - \hat{\boldsymbol{\vartheta}}_{MV}) \qquad (10A.2)$$

que es la distancia de Mahalanobis entre $\hat{\boldsymbol{\vartheta}}_0$ y $\hat{\boldsymbol{\vartheta}}_{MV}$. En efecto, si H_0 es cierta, asintóticamente $\hat{\boldsymbol{\vartheta}}_{MV}$ es centrado, con media el verdadero valor $\boldsymbol{\vartheta}_0$ y matriz de varianzas y covarianzas $[-\mathbf{H}(\hat{\boldsymbol{\vartheta}}_{MV})]^{-1}$, teniendo además una distribución normal. Como si H_0 es cierta $\hat{\boldsymbol{\vartheta}}_0$ y $\hat{\boldsymbol{\vartheta}}_{MV}$ tienen la misma esperanza, la expresión anterior puede escribirse:

$$\mathbf{w}'\mathbf{M}^{-1}\mathbf{w}$$

donde las variables normales $\mathbf{w} = \hat{\boldsymbol{\vartheta}}_0$ y $\hat{\boldsymbol{\vartheta}}_{MV}$ tienen media cero. Es claro, pues, que $2\ln\lambda$ tendrá una distribución χ^2; como el vector $\mathbf{w}$ pertenece a un espacio de dimensión r (diferencia de dimensiones entre $\hat{\boldsymbol{\vartheta}}_{MV}$ y $\hat{\boldsymbol{\vartheta}}_0$), la distribución χ^2 resultante tendrá r grados de libertad.

Apéndice 10B: Test de razón de verosimilitudes y test de multiplicadores de Lagrange

Un procedimiento alternativo de construir un contraste de hipótesis en el problema general H_0: $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$; H_1: $\boldsymbol{\vartheta} \neq \boldsymbol{\vartheta}_0$ es partir de que si H_0 es cierta la derivada de la verosimilitud en ϑ_0 debe ser próxima a cero.

Sea $\nabla\mathbf{L}(\boldsymbol{\vartheta})$ el vector columna que contiene las derivadas del soporte respecto a los componentes de $\boldsymbol{\vartheta}$. Desarrollemos $L(\boldsymbol{\vartheta})$ en serie de Taylor alrededor de $\boldsymbol{\vartheta}_0$:

$$\nabla\mathbf{L}(\boldsymbol{\vartheta}) \simeq \nabla\mathbf{L}(\boldsymbol{\vartheta}_0) + \mathbf{H}(\boldsymbol{\vartheta}_0)(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0)$$

donde $\mathbf{H}(\boldsymbol{\vartheta}_0)$ es la matriz de segundas derivadas evaluadas en $\boldsymbol{\vartheta}_0$.

Si la hipótesis $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$ es cierta, $\hat{\boldsymbol{\vartheta}}_{MV}$ debe estar próximo a $\boldsymbol{\vartheta}_0$. Por tanto, particularizando para dicho punto y teniendo en cuenta que $\nabla\mathbf{L}(\hat{\boldsymbol{\vartheta}}_{MV}) = 0$:

$$(\hat{\boldsymbol{\vartheta}}_{MV} - \boldsymbol{\vartheta}_0) \simeq [-\mathbf{H}^{-1}(\boldsymbol{\vartheta}_0)][\nabla\mathbf{L}(\boldsymbol{\vartheta}_0)] \qquad (10B.1)$$

podemos definir una medida de distancia entre $\hat{\vartheta}_{MV}$ y $\hat{\vartheta}_0$ en el mismo espíritu de (10A.2) mediante

$$(\hat{\boldsymbol{\vartheta}}_{MV} - \hat{\boldsymbol{\vartheta}}_0)'[-H(\boldsymbol{\vartheta}_0)](\hat{\boldsymbol{\vartheta}}_{MV} - \hat{\boldsymbol{\vartheta}}_0)$$

utilizando (10B.1) y aproximando la información observada por su valor esperado, la información esperada (apéndice 7B), resulta el test del gradiente o de multiplicadores de Lagrange

$$TG = [\nabla\underline{\mathbf{L}}(\vartheta_0)]'\underline{\mathbf{IE}}^{-1}(\vartheta_0)[\nabla\underline{\mathbf{L}}(\vartheta_0)] \qquad (10B.2)$$

Intuitivamente el test del gradiente mide la distancia (de Mahalanobis) entre el vector de primeras y segundas derivadas del soporte y su valor esperado en ϑ_0.

Este test es asintóticamente equivalente al de razón de verosimilitudes, pero tiene la ventaja de que no es necesario estimar $\hat{\boldsymbol{\vartheta}}_{MV}$ para realizar el test. A cambio, hay que calcular las primeras y segundas derivadas del soporte y su valor esperado en ϑ_0.

La equivalencia entre ambos procedimientos radica en que asintóticamente la verosimilitud es cuadrática y entonces ambos métodos coinciden. Por ejemplo, para un parámetro, si

$$L(\vartheta) = k(\vartheta - \hat{\vartheta}_{MV})^2$$

el test de verosimilitud es:

$$TV = 2[L(\hat{\vartheta}_{MV}) - L(\vartheta_0)] = 2[0 - k(\vartheta_0 - \hat{\vartheta}_{MV})^2] = -2k(\vartheta_0 - \hat{\vartheta}_{MV})^2$$

como:

$$\frac{dL(\vartheta)}{d\vartheta} = 2k(\vartheta - \hat{\vartheta}_{MV}) \qquad \frac{d^2L(\vartheta)}{d^2\vartheta} = 2k$$

el test del gradiente es:

$$TG = [2k(\vartheta_0 - \hat{\vartheta}_{MV})]^2/(-2k) = -2k(\vartheta_0 - \hat{\vartheta}_{MV})^2 = TV$$

11. Decisiones en incertidumbre

Abraham Wald (1902-1950)
Matemático austriaco. Creador de la teoría de análisis secuencial y uno de los fundadores de la teoría estadística de la decisión durante su trabajo dentro del Statistical Research Group en la Universidad de Columbia durante la Segunda Guerra Mundial. Se exilió a Estados Unidos en 1938 y posteriormente tomó la nacionalidad de este país.

11.1 Introducción

Para analizar las características de un problema de decisión en condiciones de incertidumbre consideremos el siguiente ejemplo: una persona tiene que optar cada mañana entre dos trayectos. La duración de cada uno depende del estado del tráfico, que, para simplificar, clasificaremos en fluido (el 10% de las veces), normal (60% de las veces) y malo (30% de los casos). Según el estado del tráfico se obtienen los tiempos de trayecto que se indican en la tabla 11.1. ¿Qué opción debe elegirse?

Este ejemplo muestra los tres componentes básicos de un problema de decisión en condiciones de incertidumbre:

1) un conjunto de opciones (a_1, ..., a_k), de las cuales debe escogerse una;

Tabla 11.1 Un problema de decisión en incertidumbre

Suceso	Probabilidad	Trayecto a_1	Trayecto a_2
$\vartheta_1 = F$	0,1	15 m.	30 m.
$\vartheta_2 = N$	0,6	35 m.	40 m.
$\vartheta_3 = M$	0,3	70 m.	50 m.

2) un conjunto de sucesos inciertos (ϑ_1, ..., ϑ_m) cuyas probabilidades supondremos conocidas;
3) una función de consecuencias, $r_{ij} = \ell(a_i \vartheta_j)$, que indica el resultado obtenido cuando se toma la acción a_i y ocurre el resultado ϑ_j. Cuando esta función mide consecuencias negativas o costes, se denomina *función de pérdida*; en el caso contrario (por ejemplo, si los resultados son ingresos monetarios) se denomina *función de beneficios*.

La decisión a tomar con estos componentes depende del *criterio de decisión*. Un criterio razonable en muchos casos es minimizar la pérdida o coste promedio o esperado (o maximizar el beneficio promedio), es decir, suponiendo una función de pérdidas:

$$\text{minimizar } CE[a_i] = \min_i \sum_{j=1}^{m} p(\vartheta_j)\, r_{ij} \tag{11.1}$$

En nuestro ejemplo, este criterio equivale a minimizar el coste esperado o tiempo promedio de trayecto, que es:

$$CE(a_1) = 0{,}1(15) + 0{,}6(35) + 0{,}3(70) = 43{,}5 \text{ minutos}$$
$$\text{CE}(a_2) = 0{,}1(30) + 0{,}6(40) + 0{,}3(50) = 42 \text{ minutos}$$

y con este criterio el trayecto elegido es el a_2, que ahorra, en promedio, un minuto y medio por trayecto.

11.2 Costes de oportunidad

Llamaremos coste de oportunidad de una opción, cuando ocurre el suceso ϑ_j, a la pérdida que se experimenta por tomar esta opción en lugar de la alternativa óptima cuando ese suceso ocurre. Por ejemplo, si tomamos el trayecto a_1 y ocurre *F*, el coste de oportunidad es cero, ya que esta opción es la mejor cuando ocurre *F*. *El coste de oportunidad es cero porque hemos*

hecho lo mejor posible. Sin embargo, el coste de oportunidad de a_2 cuando ocurre *F* es 15 minutos, la diferencia entre los 30 m. invertidos y los 15 que hubiésemos obtenido al escoger a_1.

Tabla 11.2 Costes de oportunidad de las decisiones de la tabla 11.1

Suceso	Probabilidad	$CO(a_1)$	$CO(a_2)$
F	0,1	0	15
N	0,6	0	5
M	0,3	20	0
Promedio		6	4,5

La tabla 11.2 indica los costes de oportunidad de cada acción para cada suceso de la tabla 11.1. Si ponderamos los costes de oportunidad para cada suceso por la probabilidad de este suceso, se obtiene el *coste de oportunidad esperado* de cada decisión. Se observa que la diferencia entre el coste esperado y el coste de oportunidad esperado de cada una de las dos opciones es constante. En efecto:

$$CE(a_1) - COE(a_1) = 43{,}5 - 6 = 37{,}5 = CE(a_2) - COE(a_2)$$

Este resultado es general: *la diferencia entre el coste esperado de una opción y su coste de oportunidad* (*esperado*) *es siempre constante*. La demostración de esta propiedad es simple. Supongamos que, como en el ejemplo, las consecuencias r_{ij} son pérdidas. (Si fuesen beneficios bastaría cambiar el signo.) Entonces el coste de oportunidad de la acción a_i cuando ocurre el suceso ϑ_j es:

$$CO_{ij} = r_{ij} - \min r_{ij} = r_{ij} - r_j^* \tag{11.2}$$

llamando r_j^* al coste de la mejor alternativa cuando ocurre ϑ_j. Entonces, el coste esperado será:

$$\begin{aligned} COE(a_i) &= \Sigma\, p_j\, CO_{ij} = \Sigma\, p_j\, (r_{ij} - r_j^*) \\ &= \Sigma\, p_j r_{ij} - \Sigma\, p_j r_j^* \end{aligned} \tag{11.3}$$

de donde concluimos:

$$\Sigma\, p_j r_j^* = CE(a_i) - COE(a_i) \tag{11.4}$$

El promedio del primer miembro es una constante que no depende de la acción a_i, ya que pondera el mejor resultado posible cuando ocurre cada suceso por la probabilidad de dicho suceso. En consecuencia, la diferencia entre el coste esperado de una acción y su coste de oportunidad es una constante.

Una consecuencia de esta propiedad es la equivalencia entre minimizar costes o costes de oportunidad, ya que unos se relacionan con los otros mediante una constante. Esta constante tiene una interpretación interesante que analizaremos en la sección siguiente.

11.3 El valor de la información

Consideremos de nuevo el ejemplo de la tabla 11.1. Los costes de oportunidad son debidos al estado de incertidumbre, ya que si conociésemos el suceso que va a ocurrir estos costes serían siempre cero, porque tomaríamos la opción mejor para ese suceso. En consecuencia, una forma de medir cuánto nos cuesta esta incertidumbre es calcular la diferencia entre el coste (beneficio) esperado con la mejor opción disponible y el coste (beneficio) esperado si dispusiésemos de información perfecta. Supongamos que cada día pudiéramos conocer el estado del tráfico: tomaríamos el camino a_1 cuando el tráfico fuese fluido o normal y el a_2 cuando éste fuese malo. Por tanto, el 10% de las veces circularíamos por a_1 con tráfico fluido, el 60% por a_1 con tráfico normal y el 30% por a_2 con tráfico malo. El tiempo (coste) promedio al disponer de información perfecta sería:

$$CEIP = 0{,}1 \cdot 15 + 0{,}6 \cdot 35 + 0{,}3 \cdot 50 = 37{,}5$$

Llamaremos *coste de incertidumbre o valor esperado de la información perfecta* a la diferencia entre el coste esperado con la mejor opción existente y el coste esperado con información perfecta. Como en este caso la mejor opción es a_2:

$$VEIP = CI = 42 - 37{,}5 = 4{,}5 \text{ minutos}$$

En consecuencia, podríamos ahorrar en promedio 4,5 minutos por viaje, si dispusiéramos de información perfecta.

Observemos que el *VEIP*, o coste de incertidumbre, es idéntico al coste esperado de oportunidad de la acción a_2. Esta igualdad es general: *el coste de incertidumbre es análogo al coste esperado de oportunidad de la mejor opción*. En efecto, si disponemos de información perfecta no puede existir coste de oportunidad porque siempre tomamos la mejor opción.

Para comprobar este resultado analíticamente, llamando como en la sección anterior:

$$r_j^* = \min_i r_{ij}$$

al mejor resultado cuando ocurre el suceso *j*. Entonces:

$$CEIP = \Sigma\, p_j r_j^* \tag{11.5}$$

y sustituyendo (11.5) en (11.4) concluimos:

$$\boxed{CEIP = CE(a_i) - COE(a_i)} \tag{11.6}$$

es decir, para cualquier opción, la diferencia entre su coste y su coste de oportunidad es el coste esperado con información perfecta, que es único para el problema de decisión. En particular, esta relación sigue siendo válida para la opción con información perfecta, ya que para ella el *COE* se anula.

Supongamos ahora que a_i es la opción óptima. Entonces, el valor esperado de información perfecta o coste de incertidumbre es, utilizando (11.6):

$$\boxed{VEIP = CE(a_i) - CEIP = COE(a_i)} \tag{11.7}$$

Por tanto:

1) La diferencia entre el coste esperado de una opción y su pérdida esperada de oportunidad es constante e igual al coste esperado con información perfecta.
2) El valor (esperado) de la información perfecta es el coste de oportunidad (esperado de la acción óptima).

Cuando los resultados son positivos (beneficios en lugar de costes) el razonamiento es análogo. Entonces r_j^* es el máximo de las filas y

$$VEIP = BEIP - BE(a_i) = COE(a_i) \tag{11.8}$$

donde *B* representa beneficios en lugar de costes.

11.4 Decisiones con información muestral

Es frecuente que en un problema de decisión podamos reducir la incertidumbre recogiendo información mediante una muestra o realizando un experimento. En general inicialmente la incertidumbre está reflejada por una distribución de probabilidad que supondremos discreta, y representaremos las probabilidades iniciales por $P(\vartheta_j)$, donde ϑ_j puede ser un suceso o un parámetro. Entonces, después de observar la muestra M las probabilidades se modificarán mediante el teorema de Bayes:

$$P(\vartheta_j|M) = \frac{P(M|\vartheta_j)P(\vartheta_j)}{P(M)} \tag{11.9}$$

donde $P(\vartheta_j|M)$ es la probabilidad a posteriori del suceso ϑ_j cuando se ha observado la muestra M. El denominador:

$$P(M) = \sum_{j=1}^{m} P(M|\vartheta_j)P(\vartheta_j) \tag{11.10}$$

representa la probabilidad de obtener la muestra M. Ahora la decisión óptima será la que conduzca a un coste esperado menor (beneficio mayor) con la información disponible, es decir:

$$\min_i CE(a_i) = \Sigma P(\vartheta_j|M)r_{ij} \tag{11.11}$$

que es similar a (11.1) pero con las probabilidades a priori, $P(\vartheta_j)$, reemplazadas por las probabilidades a posteriori, $P(\vartheta_j|M)$.

11.4.1 El valor de la muestra

Una de las ventajas principales de la teoría de la decisión es que permite evaluar el valor de la información *antes* de tenerla. Esta evaluación permite decidir si es rentable o no económicamente disponer de ella. En efecto, antes de disponer de la muestra podemos calcular los posibles resultados, M_1, ..., M_T y sus probabilidades relativas con la información disponible a priori, $P(M_1)$, ..., $P(M_T)$ mediante (11.10). A continuación, podemos calcular para cada M_i la mejor opción posible, lo que supone calcular las probabilidades $P(\vartheta_j|M_i)$ mediante (11.9) y aplicar (11.11) para obtener la mejor opción. Sea $CE(M_i)$ el coste esperado de la opción más favorable cuando se da el resultado M_i. Entonces, el coste esperado evaluado *antes* de tomar la muestra, pero contando con los posibles resultados de ésta, es:

$$CEIM = \sum_{i=1}^{T} P(M_i)\ \mathrm{CE}(M_i) \tag{11.12}$$

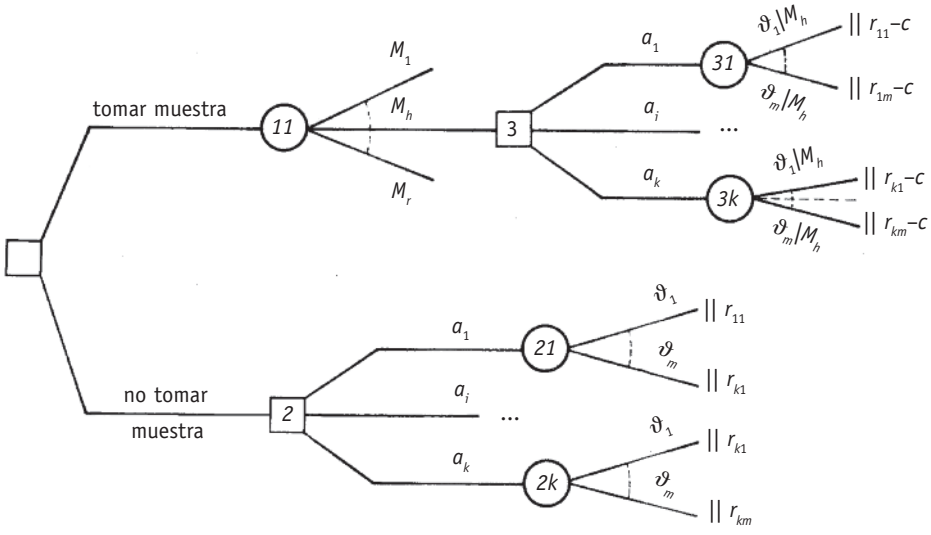
Si a_i es la mejor opción sin tomar la muestra y su coste en $CE(a_i)$, el valor esperado de la información muestral (*VEIM*) es:

$$VEIM = CE(a_i) - CEIM \tag{11.13}$$

que tiene la misma estructura que (11.5).

Este análisis se resume en el árbol de decisión de la figura 11.1. Suponemos que existen k posibles acciones, $(a_1, ..., a_k)$, que las consecuencias de cada acción dependen de un conjunto de m sucesos inciertos, $(\vartheta_1, ... \vartheta_m)$ con probabilidades conocidas y que los resultados $r_{ij} = \ell(a_i \vartheta_j)$, donde $i = 1, ..., k$ y $j = 1, ..., m$ son también conocidos. En este gráfico el símbolo □ indica un punto de decisión y el ○ un punto aleatorio, entendiendo por ello que el camino a partir de ese punto viene determinado por el azar. La primera decisión es tomar o no la muestra. Si no la tomamos, seguimos por la rama inferior y nos encontramos con el punto de decisión indicado por 2. La decisión óptima en este punto como en todos los puntos de decisión es minimizar el coste esperado. Calcularemos para cada posible decisión a_i el coste esperado multiplicando los resultados r_{ij} por las probabilidades de obtenerlos, $P(\vartheta_i)$.

Figura 11.1 Árbol de decisión con información muestral



Si tomamos la muestra nos encontraremos con el punto incierto ⑪, ya que el resultado de la muestra es desconocido. Para cada posible valor muestra, M_1, ..., M_T, tendremos que tomar después una decisión y el resultado dependerá de cuál de los sucesos inciertos (ϑ_1, ... ϑ_m) ocurre. Supongamos que la muestra proporciona el resultado M_h. Entonces, en lugar de utilizar las probabilidades a priori, $P(\vartheta_i)$, utilizaremos las probabilidades a posteriori, $P(\vartheta_i/M)$, que se calcularán por el teorema de Bayes. Las consecuencias finales también variarán, porque tendremos que restar a los resultados r_{ij} el coste de tomar la muestra, c. Por lo demás el análisis será similar: en cada punto de incertidumbre calcularemos el valor esperado, y en cada punto de decisión tomaremos la acción que lleve a un coste esperado menor, o a un beneficio esperado mayor.

En resumen, el árbol de decisión pone de manifiesto los dos tipos de estructuras posibles en un problema de decisión en incertidumbre. La primera son los puntos aleatorios, donde el resultado depende de causas que no controlamos, pero cuyas probabilidades suponemos conocidas. Los puntos de decisión se resumen en su valor esperado, promediando las consecuencias con las probabilidades. La segunda son los puntos de decisión donde podemos escoger el camino a seguir. En esos puntos tomaremos la acción que lleve a un menor coste esperado o, lo que es equivalente, a un mayor beneficio esperado.

Cuando exista información muestral, el análisis del árbol de decisión puede resumirse en los pasos siguientes:

1) Comenzar con los nudos aleatorios terminales (en la figura 11.1 los nudos [*31* a *3k*] en el caso de tomar la muestra y [*21 a 2k*] si no la tomamos).
2) En los nudos de decisión ② y ③ escoger la alternativa óptima (mayor beneficio esperado) y tomar ese valor como resultado del punto de decisión.
3) Calcular el valor esperado en el nudo aleatorio 11.
4) Comparar los dos valores esperados en el punto de decisión ① resultantes de tomar o no tomar muestra y escoger aquel que conduzca a un valor esperado más alto.

Este algoritmo de *promediar y retroceder* puede siempre aplicarse sea cual sea la complejidad del problema. Como ilustramos en el ejemplo 11.1 y en la figura 11.2, cuando existan decisiones secuenciales el árbol de decisión puede ser más complicado, pero siempre podemos aplicar el algoritmo general siguiente:

1. Calcular todas las probabilidades a posteriori necesarias mediante el teorema de Bayes y las consecuencias e introducirlas en el árbol.

2. Comenzar con los nudos aleatorios finales y sustituir cada nudo por su valor esperado.
3. Tomar en cada punto de decisión la acción que lleve a un mayor valor esperado.
4. Continuar desde el final hasta el principio utilizando las reglas 2 y 3 hasta determinar la primera acción a tomar y la secuencia consiguiente de decisiones.

Estos principios se ilustran en el ejemplo 11.1.

Ejemplo 11.1

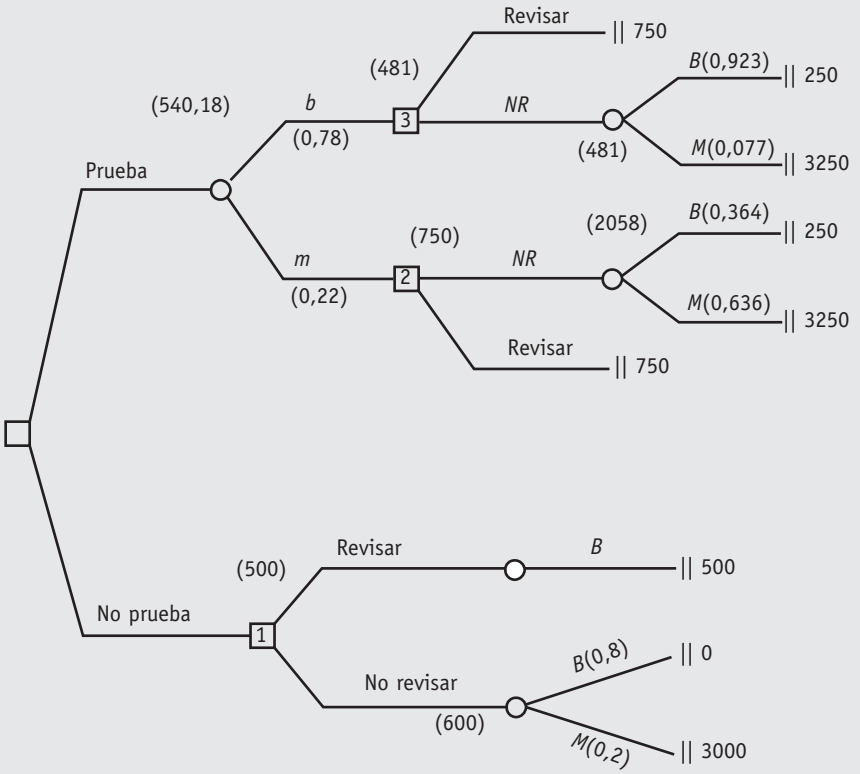
Una empresa se plantea la opción de revisar (*R*) o no (*N R*) un proceso antes de comenzar la actividad de cada día. El coste de revisar el proceso es de 500 euros, pero garantizamos que el proceso funcionará satisfactoriamente todo el día. Si no revisamos el proceso no incurrimos en ningún coste si no hay problemas, pero si se producen problemas el coste de parar y arreglarlo es de 3000 euros. Se conoce que la probabilidad de que se produzcan problemas cuando el proceso no se ha revisado es de 0,2, mientras que la revisión garantiza la ausencia de problemas durante el día. Existe la posibilidad de hacer cada mañana una prueba rápida que cuesta 250 euros y nos puede indicar si el proceso requiere o no revisión. Si decidimos hacer la prueba podemos decidir con más información si revisamos o no en función del resultado de la prueba, ya que, aunque la prueba es informativa, tiene cierto margen de error. En concreto se conoce que *P* (prueba que indique que el proceso está bien / proceso está bien) = 0,9 y *P* (prueba que indique que el proceso necesita revisión / el proceso necesita revisión) = 0,7.

La estructura del problema se presenta en la figura 11.2.

La primera opción a tomar es decidir si realizamos o no la prueba. Si no la hacemos (rama inferior) tendremos que decidir si revisar o no revisar antes de comenzar. Si revisamos, sabemos que el proceso funcionará bien, lo que representamos en el árbol mediante el suceso *B*, y el coste de esta acción es 500 euros. Si no revisamos, el coste depende de que el proceso esté bien, suceso *B*, o que esté mal, suceso *M*. En el primer caso el coste es cero. En el segundo, incurriremos en el coste de comenzar y parar después para revisar, cuando se detecte el problema, lo que supone un coste de 3000 euros.

Si realizamos la prueba, los costes dependen de su resultado. Llamemos *b* al suceso: la prueba indica que el proceso está bien, y *m* al suceso: la prueba indica que el proceso está mal. En función del resultado podemos

Figura 11.2 Análisis del problema con un árbol de decisión



plantearnos de nuevo si revisar o no revisar, pero ahora las probabilidades de los sucesos *B* y *M* se calculan con el teoroma de Bayes. La probabilidad de que el proceso funcione bien cuando la prueba indica esto será:

$$P\,(B \mid b) = \frac{P\,(b \mid B)\; p\;(B)}{P\;(b)}$$

y

$$P(b) = P(b/B)P(B) + P(b/M)P(M).$$

Según los datos del problema $P(M) = 0{,}2$ y, en consecuencia, $P(B) = 1 - 0{,}2 = 0{,}8$. Por otro lado, $P(b/B) = 0{,}9$, y $P(m/M) = 0{,}7$, lo que supone $P(b/M) = 1 - 0,\ 7 = 0{,}3$. Con estos datos calculamos

$$P(b) = 0{,}9 \times 0{,}8 + 0{,}3 \times 0{,}2 = 0{,}78$$

con lo que obtenemos

$$P(B \mid b) = \frac{0{,}9 \cdot 0{,}8}{0{,}78} = 0{,}923$$

$$P(M \mid b) = 1 - 0{,}923 = 0{,}077$$

Análogamente, la prueba indicará que el proceso está mal (m) con probabilidad:

$$P(m) = P(m \mid M)P(M) + P(m \mid B)P(B) = 0{,}7 \times 0{,}2 + 0{,}1 \times 0{,}8 = 0{,}22$$

que podemos calcular también más simplemente por diferencia:

$$P(m) = 1 - P(b) = 0{,}22.$$

Las probabilidades a posteriori en este caso serán:

$$P(M \mid m) = \frac{P(m \mid M)P(M)}{P(m)} = \frac{0{,}7 \times 0{,}2}{0{,}22} = 0{,}636$$

$$P(B \mid m) = 1 - 0{,}636 = 0{,}364$$

Estas probabilidades se han llevado al árbol de la figura 11.2. Para completar la estructura del problema debemos introducir las consecuencias, en este caso los costes, de las decisiones. Si hacemos la prueba y revisamos, el coste es 750 euros sin incertidumbre. Si decidimos no revisar después de la prueba, y el proceso está bien, tendremos sólo el coste de la prueba, 250 euros, mientras que si el proceso está mal incurriremos en un coste de 3.000 + 250 = 3.250 euros. Estos valores se indican en el árbol.

Una vez completada la estructura del problema, podemos aplicar el algoritmo de promediar y retroceder. Comenzamos con el punto aleatorio resultante de las acciones: no prueba no revisar, en la parte inferior del árbol. El valor esperado en ese punto es

$$CE(N\ R) = 0,\ 8 \cdot 0 + 0{,}2 \cdot (3.000) = 600 \text{ euros}$$

Si retrocedemos ahora al punto de decisión 1 tenemos que comparar la opción revisar, con coste 500, con la de no revisar, con coste esperado de 600, por lo que decidiremos revisar. Llevaremos ese valor de 500 al punto de decisión y ya hemos terminado esa rama: la decisión es revisar y el coste esperado 500.

Pasamos ahora a evaluar otro de los nudos aleatorios terminales. Continuando de abajo arriba nos encontramos el nudo terminal definido por la

secuencia: Prueba, *m*, *NR*, donde podemos obtener unos costes de 250 o 3.250 con probabilidades 0,364 o 0,636. El valor esperado es

$$CE(NR/m) = 0,364 \cdot 250 + 0,636 \cdot (3.250) = 2.058 \text{ euros}$$

Si comparamos en el nudo de decisión 2 las dos alternativas, *NR* lleva a 2058, mientras que revisar lleva a 750, con lo que decidiremos revisar y el coste de la mejor opción se coloca encima del punto de decisión.

El siguiente punto aleatorio es el definido por Prueba, *b*, *NR*, y el valor esperado es

$$CE(NR/b) = 0,923 \cdot 250 + 0,077 \cdot (3.250) = 481 \text{ euros}$$

y si comparamos ahora este valor con las 750 de revisar, es claro que cuando la prueba indica que el proceso está bien es mejor no revisar, con un coste esperado de 481. Con esto podemos retroceder al último punto aleatorio que define el resultado de la prueba. Nos encontramos que si la prueba indica *b*, lo que ocurrirá con probabilidad 0,78, el coste esperado es de 481, mientras que si la prueba indica *m*, lo que ocurrirá con probabilidad 0,22, el coste esperado es de 750. El coste esperado en el nudo aleatorio es:

$$CE \text{ (Prueba)} = 0,78 \cdot 481 + 0,22 \cdot (750) = 540,18 \text{ euros}$$

A continuación retrocedemos al punto inicial de decisión, donde tenemos dos alternativas: hacer la prueba, que conduce a una secuencia de decisiones que produce un coste esperado de 540,18, o no hacerla, que conduce a un coste esperado de 500. En consecuencia, es mejor no realizar la prueba.

Como ejercicio vamos a calcular el coste de la información perfecta, que es el coste de oportunidad de la mejor acción. Los costes de oportunidad cuando no se realiza la prueba son:

$$CO(R) = 500(0,8) + 0(0,2) = 400 \text{ euros}$$

$$CO(NR) = 0 \cdot (0,8) + 2.500(0,2) = 500 \text{ euros}$$

y el coste de incertidumbre, que es el coste de oportunidad de la mejor opción, *R*, es, por tanto, igual a 400 euros. Ésta es la cantidad máxima que podemos pagar por cualquier información. Como la prueba vale menos que el coste de incertidumbre, puede valer la pena analizarla. Si su coste fuese mayor de 400 euros quedaría automáticamente descartada. Comprobemos también que la estrategia con menor coste debe tener también un

menor coste de oportunidad. La diferencia entre coste esperado y coste de oportunidad:

$$CE(R) - CO(R) = CE(NR) - CO(NR) = 100 \text{ euros}$$

es el coste esperado con información perfecta (*CEIP*) . En efecto, si dispusiésemos de información exacta del estado del proceso sólo revisaríamos cuando el proceso lo necesite, lo que ocurrirá con probabilidad 0,2, obteniendo un coste de:

$$CEIP = 0{,}2 \cdot (500) = 100 \text{ euros}$$

Para evaluar si vale la pena hacer la prueba, observemos que los resultados al hacerla, sin incluir el coste de la prueba, son:

a) Con probabilidad 0,22 la prueba resulta en (*m*) y decidimos revisar con un coste de 500 euros (750 que aparecen en el árbol menos los 250 del coste de la prueba).
b) Con probabilidad 0,78 la prueba resulta en (*b*) y decidimos no revisar con un coste de 231 (481 menos 250).

Por tanto, el coste esperado con la información de la prueba (muestral) es:

$$CEIM = 0{,}22 \cdot (500) + 0{,}78 \cdot (231) = 290{,}18 \text{ euros.}$$

Como el coste con la mejor estrategia sin realizar la prueba es 500 y al realizarla 290,18, la diferencia entra ambas cantidades será el valor esperado de la información muestral:

$$VEIM = CE(R) - CEIM = 500 - 290{,}18 = 209{,}82 \text{ euros.}$$

Si el coste de realizar la prueba es menor que el valor esperado de la información muestral, valdrá la pena hacerla. Como el coste es 250, mayor que su valor esperado de 209,82, no conviene realizar la prueba.

11.5 Utilidad

11.5.1 El criterio del valor esperado

El criterio de la esperanza matemática o del valor esperado establece que entre dos opciones cuyas consecuencias están medidas en unidades homo-

géneas es preferida aquella que conduzca a un valor esperado mayor. Este criterio es razonable cuando: (1) la decisión es repetitiva; (2) las consecuencias no son muy importantes para el decisor. Por ejemplo, consideremos las dos opciones de la tabla 11.3; ¿cuál parece preferida?

Tabla 11.3

Suceso	Probabilidad	Opción A	Opción B
cara	0,5	35,0 E	12,0 E
cruz	0,5	–10,0 E	8,0 E

De acuerdo con el criterio de la esperanza matemática (*EM*), la opción *A* que tiene una esperanza de beneficio de 12,5 euros es preferida a la *B*, donde la esperanza es sólo 10,0 euros. Este criterio es razonable si la decisión fuese repetitiva, ya que, a largo plazo, la opción *A* proporcionará, con probabilidad que tiende a uno, beneficios superiores a la opción *B*. Por ejemplo, si *A* y *B* representan dos tipos de créditos que un banco puede ofrecer, el banco ganará, a largo plazo, un 25% más con el *A* que con el *B*.

Sin embargo, para una persona que va a elegir sólo una vez, el criterio de la esperanza matemática es discutible, y muchas personas prefieren *B* a *A*. El problema radica en que el valor de una cantidad monetaria es distinta para distintos decisores. Por ejemplo, con el criterio del valor esperado la opción *A* sigue siendo mejor que la *B* si multiplicamos o dividimos todas las consecuencias por 100, lo que probablemente no es cierto para el lector. Un criterio general de decisión debe tener en cuenta estos factores.

11.5.2 El riesgómetro

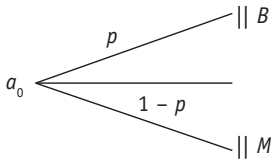
Comparamos temperaturas con un termómetro, longitudes con un metro o resistencias con un voltímetro. Para comparar consecuencias en situaciones de riesgo vamos a introducir un instrumento de medida que llamaremos *riesgómetro*. Consideremos un problema de decisión con distintas opciones $\{a_1, ..., a_k\}$ y consecuencias $\{r_{ij}\}$ que dependen de ciertas probablidades p_{ij}. Supondremos que el decisor acepta los siguientes axiomas de coherencia:

1. *Ordenación*. Todas las consecuencias son comparables, y dadas dos cualesquiera, r_A y r_B, el decisor puede siempre decir si prefiere r_A a r_B (que escribiremos $r_A > r_B$), r_B a r_A ($r_B > r_A$) o está indiferente entre ambas ($r_A = r_B$). En este último caso diremos que las consecuencias son equivalentes.

2. *Transitividad*. Si $r_A > r_B$ y $r_B > r_C$ entonces $r_A > r_C$. Análogamente, si $r_A = r_B$ y $r_B = r_C$, entonces $r_A = r_C$.
3. *Sustitución*. Las preferencias del decisor entre dos opciones inciertas no se modifican al sustituir una consecuencia por otra equivalente. En consecuencia, dadas dos opciones a_1 y a_2 que sólo difieren en una consecuencia:

donde $\Sigma p_i = 1$, si $r_{A'}$ es equivalente a r_A, entonces el decisor debe estar indiferente entre las opciones a_1 y a_2.

Estos tres axiomas permiten construir una escala de preferencias ante el riesgo para cada persona. Suponga el lector que es el decisor y que la *mejor* consecuencia posible en el problema de decisión planteado es *B y M* la *peor*. Vamos a llamar *riesgómetro* para ese problema a una opción del tipo:



donde existe una probabilidad p de obtener B y $1 - p$ de obtener M. En esta opción B y M son siempre fijos y variando p podemos calibrar por comparación nuestras preferencias por las consecuencias de la forma siguiente:

1) ordenemos todas las consecuencias posibles, en orden ascendente de preferencia:

$$M \leq r_1 \leq r_2 \ldots \leq r_h \leq B$$

donde ≤ indica preferido o equivalente;
2) planteamos un problema de decisión simple donde podemos optar entre obtener r_i con certeza o la opción del riesgómetro con proba-

bilidad p_i de obtener B y nos preguntamos qué valor de p_i hace que ambas opciones sean equivalentes para nosotros. Es decir, la elección es entre a_i y a_0.

y se trata de fijar p_i para que exista indiferencia entre ambas opciones. Es claro que si p_i es próximo a 1, a_0 será preferida, y si p_i es próximo a cero, a_i será preferida. En efecto, si $p_i = 1$, el riesgómetro equivale a obtener B con certeza, y si $p_i = 0$, a obtener M con certeza. Entonces existirá un valor p_i que haga ambas alternativas equivalentes. Llamaremos *utilidad* de r_i a este valor, y escribiremos:

$$u(r_i) = p_i$$

y se verificará, $u(B) = 1$, $u(M) = 0$ y:

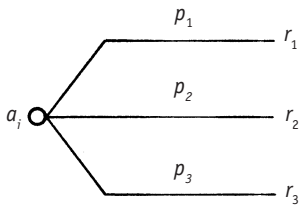
$$0 \leq u(r_i) \leq 1$$

En resumen, este procedimiento asigna a cada consecuencia un número que llamaremos utilidad que tiene las propiedades de una probabilidad. De esta manera podemos asignar a todas las consecuencias del problema de decisión un número entre 0 y 1 que verifica: si $r_i \leq r_j$; $u(r_i) \leq u(r_j)$.

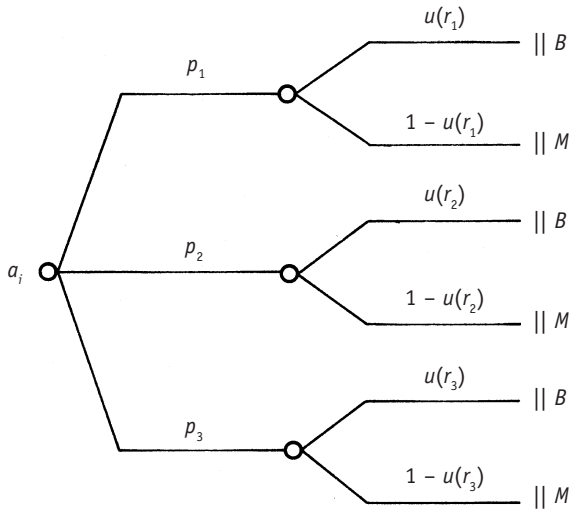
11.5.3 La función de utilidad

Consideremos una opción cualquiera en el problema de decisión. Vamos a demostrar que si sustituimos las consecuencias por sus utilidades calculadas por el método del riesgómetro, la opción preferida debe ser la de mayor utilidad esperada.

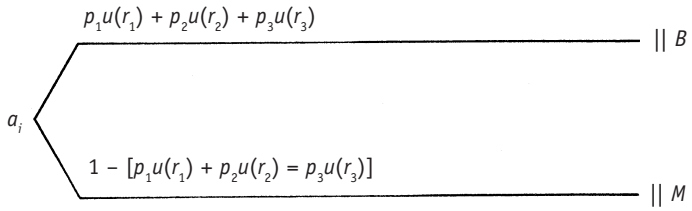
En efecto, cualquier opción a_i, por ejemplo:



es equivalente, de acuerdo con el principio de sustitución a:



donde hemos sustituido cada resultado por el riesgómetro equivalente. En definitiva, esta alternativa se reduce de nuevo a un riesgómetro, ya que las consecuencias finales son únicamente B y M y a_i puede escribirse:



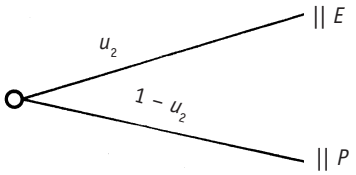
y es un riesgómetro con probabilidad $\Sigma p_i u(r_i)$. De acuerdo con el criterio establecido, la utilidad de a_i será la probabilidad de obtener B, que resulta ser:

$$u(a_i) = \Sigma p_i u(r_i)$$

Si repetimos este análisis para cada una de las opciones, al final reducimos todas ellas a riesgómetros con distintas probabilidades de obtener B. En consecuencia, la opción preferida será aquella para la cual esta probabilidad es máxima (mayor utilidad).

Este análisis conduce al siguiente principio general: *si sustituimos las consecuencias por sus utilidades*, *la mejor opción es la de mayor utilidad esperada*.

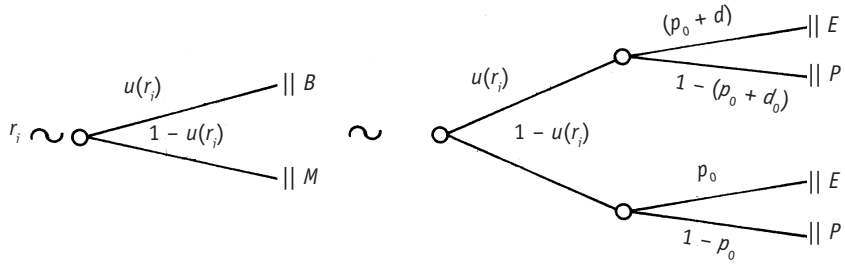
Es importante comprobar que las decisiones obtenidas con el principio de maximizar la utilidad esperada no dependen de las consecuencias B y M elegidas para construir el riesgómetro. En efecto, supongamos que tomamos otras consecuencias arbitrarias $E > B$ y $P < M$ como referencia. Entonces, llamando u_2 a la probabilidad de ganar en el nuevo riesgómetro:



podemos trasladar los valores $u(r_i)$ anteriores a esta nueva escala decidiendo la probabilidad que hace indeferente este riesgómetro $[(u_2(E); E, P]$ a recibir M con certeza. Sea p_0 este valor, entonces $u_2(M) = p_0$. Repitiendo esta cuestión para B obtenemos la equivalencia de B en la nueva escala, supongamos que:

$$u_2(B) = p_0 + d_0$$

donde, como B es preferido a M, d_0 es positivo. Entonces, la utilidad $u_2(r_i)$ en esta nueva escala de una consecuencia con utilidad $pi = u(r_i)$ en la antigua escala definida por $[u(r_i); B, M]$ se obtendrá sustituyendo B y M por sus utilidades en la nueva escala, con lo que el decisor está indiferente entre:



lo que implica una probabilidad de obtener E, que es la utilidad en la nueva escala, de:

$$\begin{aligned} u_2(r_i) &= u(r_i)(p_0 + d_0) + [1 - u(r_i)]p_0 \\ &= p_0 + d_0 u(r_i) \end{aligned}$$

Este análisis muestra que un cambio de escala equivale a una transformación lineal de las utilidades. Por tanto el *orden* de elección entre alternativas no se verá afectado por el cambio de escala, y podemos aplicar una transformación lineal arbitraria a las utilidades sin afectar a sus propiedades. También indica que al comparar utilidades interesa únicamente las diferencias relativas:

$$\frac{u_2(r_i) - u_2(r_j)}{u_2(r_j)} = \frac{u(r_i) - u(r_j)}{u(r_j)} = \text{cte}$$

ya que serán invariantes para cualquier transformación lineal.

11.6 La curva de utilidad monetaria

El método del riesgómetro puede aplicarse siempre con independencia de cómo se midan las consecuencias. Cuando éstas sean unidades monetarias homogéneas, un método más efectivo que calibrar su utilidad una a una es decidir globalmente la forma de la curva $u(x)$ que proporciona la utilidad de cualquier valor. Este procedimiento es análogo a ajustar un modelo de distribución de probabilidad en lugar de las probabilidades de los sucesos individuales. La función de utilidad debe ser creciente, ya que si x_1 y x_2 son cantidades monetarias:

$$\text{si } x_2 > x_1 \Rightarrow u(x_2) > u(x_1)$$

por tanto, $u(x)$ tiene propiedades análogas a una función de distribución. El equivalente de la función de densidad es su derivada, $u'(x)$, que es siempre positiva. Esta función, como la de verosimilitud, puede modificarse arbitrariamente al cambiar la escala de medida, lo que sugiere estudiar su logaritmo. El equivalente de la tasa de discriminación es la aversión local al riesgo:

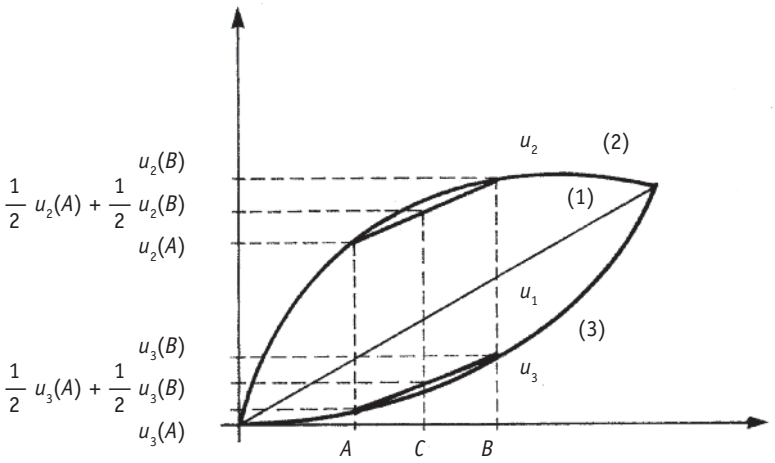
$$r(x) = -\frac{d \ln u'(x)}{dx} = -\frac{u''(x)}{u'(x)} \qquad (11.14)$$

que, es fácil demostrar, caracteriza completamente la función de utilidad.

La figura 11.3 presenta los tres tipos básicos de funciones de utilidad. En el caso (1) la función de utilidad es una recta, $u(x) = x$, y equivale a decidir con el valor esperado. En efecto la utilidad de una opción *H:* ($^1/_2$, $^1/_2$; *A*, *B*) es:

$$u_1(H) = 0{,}5\ u_1(A) + 0{,}5\ u_1(B) = (A + \text{B})/2 = C$$

Figura 11.3 Funciones de utilidad: (2) aversión al riesgo; (1) neutral y (3) propensión al riesgo



que es el centro del intervalo (*A*, *B*). En el caso (2) la función de utilidad tiene segunda derivada negativa, lo que indica que incrementos constantes monetarios producen incrementos decrecientes de utilidad. La utilidad de la opción *H* anterior es ahora:

$$u_2(H) = 0{,}5\ u_2(A) + 0{,}5\ u_2(B) < u_2(C)$$

ya que al estar la función por encima de la recta que une los dos puntos la utilidad de *H*, que es el promedio de *u*(*A*) y *u*(*B*), es siempre menor que la utilidad de *C*, promedio de *A* y *B*.

La cantidad:

$$u_2(H) - u_2(C)$$

se denomina *prima de riesgo:* es la diferencia en utilidad para el decisor entre la utilidad de una opción incierta y la utilidad de una cantidad segura igual a la esperanza matemática de esta opción. Las funciones de utilidad del tipo (2) tienen siempre una prima de riesgo positiva y describen *aversión al riesgo*.

El caso (3) es el opuesto: la segunda derivada de la función es positiva, indicando incrementos de utilidad crecientes por incrementos monetarios constantes. En este caso, la utilidad de *H* será:

$$u_3(H) = 0{,}5\ u_3(A) + 0{,}5\ u_3(B) > u_3(C)$$

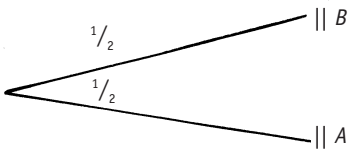
ya que la curva va por debajo de la recta: la prima de riesgo es negativa y existe *propensión al riesgo*.

Dentro de estos comportamientos la evolución de la prima de riesgo, que puede demostrarse viene determinada por $r(x)$, define la forma de la curva. Por ejemplo, en decisores con aversión al riesgo, la aversión puede ser constante o aumentar o disminuir con las cantidades involucradas. Se demuestra en el ejercicio 11.7 que si $r(x)$ es constante y positiva, la función de utilidad es exponencial con ecuación:

$$u(x) = (1 - e^{-rx})/r \tag{11.15}$$

donde x representa la riqueza del individuo y r es la aversión al riesgo (cuando menor es r, menor es la aversión al riesgo). Cuando $r \to 0$, esta función tiende a $u(x) = x$, es decir, a la función de utilidad lineal.

La utilidad exponencial tiene la ventaja de ser simple, fácil de ajustar y representar una buena aproximación a otras funciones más complicadas. El único parámetro desconocido es r, que puede determinarse como sigue: consideremos una opción simple:



y sea C la cantidad cierta equivalente a esta opción; entonces, el valor de C debe verificar:

$$(1 - e^{-rC}) = {}^1/_2\,(1 - e^{-rA}) + {}^1/_2\,(1 - e^{-rB})$$

$$C = -\,[\ln({}^1/_2\,e^{-rA} + {}^1/_2\,e^{-rB})]/r \tag{11.16}$$

La ecuación (11.16) permite obtener r por el siguiente método aproximado: comenzamos suponiendo un valor r_0 inicial y calculamos un valor C_0 con (11.16). Si $C < C_0$, la valoración dada es más alta que la de una curva con r_0, por lo que la aversión al riesgo es menor y hay que reducir r. Tomamos $r_1 < r_0$ y repetimos el cálculo para obtener C_1 y así sucesivamente. Si el valor calculado con r_j en (11.16) es $C_j < C_0$, habrá que aumentar r y utilizar $r_{j+1} > r_j$ en la siguiente iteración.

Ejemplo 11.2

Una librería debe decidir cuántas unidades pedir de un libro de texto. Suponemos para simplificar que los pedidos se hacen por paquetes de 20 libros. Los libros no vendidos suponen un coste unitario de 10 euros por gastos financieros y de devolución, mientras que los vendidos proporcionan un beneficio unitario de 40 euros. Por otro lado, hay un descuento a partir de 75 libros que hace que el beneficio unitario a partir de 75 sea de 50 euros. Las probabilidades de la demanda, estimadas por datos históricos, se dan en la tabla.

Demanda	Probabilidad
40	0,2
60	0,4
80	0,3
100	0,1

Analizar este problema con utilidad lineal y exponencial si la valoración de la opción ($^1/_2$, $^1/_2$, 2.000, 1.000) por el librero es de 1.300 euros.

Las opciones posibles son ordenar 40, 60 80 o 100 libros. Las consecuencias de cada opción se indican en la tabla, en euros.

Demanda	Prob.	*a*(40)	*a*(60)	*a*(80)	a(100)
40	0,2	1.600	1.400	1.200	1.000
60	0,4	1.600	2.400	2.200	2.000
80	0,3	1.600	2.400	4.000	3.800
100	0,1	1.600	2.400	4.000	5.000

Las consecuencias serán el beneficio obtenido restando a los beneficios de las ventas los costes de los libros no vendidos. Por ejemplo, la consecuencia de ordenar 80, a(80) y que la demanda sea de 40 se ha calculado así:

$$40 \times 40 - 40 \times 10 = 1.200 \text{ euros}$$

El valor esperado de cada opción es:

$$EM(40) = 1.600; \quad EM(60) = 1.400 \times 0{,}2 + 2.400 \times 0{,}8 = 2.200$$
$$EM(80) = 2.720; \quad EM(100) = 2.640$$

por lo que, con el valor esperado (utilidad lineal), la mejor opción es $a(80)$.

Supongamos ahora una función de utilidad exponencial. La fórmula (11.16) proporciona las siguientes valoraciones para una opción ($^1/_2$, $^1/_2$; 2.000, 1.000) en función de r:

C. resultante	**1.438**	**1.380**	**1.328**	**1.283**	**1.246**	**1.215**	**1.169**
r	0,0005	0,001	0,0015	0,002	0,0025	0,0030	0,004

en consecuencia, el valor de r es del orden de 0,0018. Tomando este valor en (11.15) se calculan las utilidades de las distintas opciones, que se presentan en la siguiente tabla conjuntamente con las utilidades esperadas.

Demanda	**Prob.**	***a*(40)**	***a*(60)**	***a*(80)**	**a(100)**
40	0,2	5.244	5.108	4.915	4.637
60	0,4	5.244	5.482	5.450	5.404
80	0,3	5.244	5.482	5.551	5.549
100	0,1	5.244	5.482	5.551	5.555
Esperada	—	5.244	5.407	5.383	5.309

Se observa que, ahora, la mejor opción es ordenar 60, que tiene la mayor utilidad esperada (5.407). El valor monetario equivalente de esta utilidad es:

$$5.407 = \frac{1}{0{,}18}\,(1 - e^{-0{,}18x})$$

$$x = 2.012 \text{ euros}$$

y el efecto de la aversión al riesgo ha sido descontar las cantidades monetarias en función del riesgo involucrado. La alternativa $a(60)$ ha pasado de valer 2.200, de acuerdo con el valor esperado, a 2.012, de acuerdo con la utilidad esperada.

Ejercicios 11

11.1. Una empresa petrolera tiene que decidir si perforar o no en una zona. Los resultados y las probabilidades se indican en la tabla. Calcular la mejor decisión con el criterio de la esperanza matemática.

Petróleo	Probabilidad	Perforar	No perforar
nada (N)	0,5	–20	0
poco (P)	0,3	+10	0
mucho (M)	0,2	+100	0

11.2. Calcular el valor esperado de la información perfecta en el problema anterior.

11.3. ¿Cuánto podría pagarse por una exploración sísmica que sólo puede dar como resultado *n* (no existe petróleo) o *s* (existe petróleo) si su fiabilidad es del 80% (es decir, P[expl = no(n)|N] = 0,8; P[expl = si(s)|Petróleo] = 0,8)?

11.4. Suponiendo utilidad exponencial, con $r = 0{,}2$, calcular qué opción es preferida en 11.1.

11.5. Analizar el problema de la tabla 11.1 de los tiempos de transporte con utilidad exponencial $r = 0{,}1$ tomando los tiempos como negativos para que representen consecuencias positivas.

11.6. Demostrar que si la aversión al riesgo es constante, la utilidad es exponencial.

11.7. Demostrar que la función de utilidad $u(x) = \ln(x + a)$ tiene aversión decreciente al riesgo.

11.7 Inferencia y decisión

11.7.1 Estimación y decisión

Cualquier problema de estimación de un parámetro ϑ puede verse como un caso particular de decisión donde el conjunto de acciones coincide con el conjunto de sucesos: ambos son iguales al conjunto de valores posibles del parámetro. En efecto, cada acción es del tipo «tomar como estimación $\hat{\vartheta}_i$» y existen tantas opciones posibles como valores pueda tener el parámetro. En la estimación clásica no existe una distribución de probabilidad sobre los valores de ϑ, por lo que la solución no es directa. Es costumbre definir

la función de consecuencias como función de pérdida de oportunidad, $\ell(\hat{\vartheta}, \vartheta)$, que toma el valor cero si $\hat{\vartheta} = \vartheta$. Entonces, dado un estimador $\hat{\vartheta}$ —una regla de decisión $\hat{\vartheta}(\mathbf{X})$, siendo $\mathbf{X}$ la muestra—, el valor esperado de la pérdida de oportunidad se denomina *riesgo* del estimador, y viene dado por:

$$R(\hat{\vartheta}, \vartheta) = \int \ell(\vartheta, \hat{\vartheta}) f(\mathbf{X} \mid \vartheta)\, d\mathbf{X}$$

donde $\mathbf{X}$ es la muestra. Por ejemplo, si tomamos:

$$\ell(\vartheta, \hat{\vartheta}) = k(\vartheta - \hat{\vartheta})^2$$

el riesgo de un estimador equivale a su error cuadrático medio. La decisión óptima (el estimador óptimo) será aquel con riesgo menor para todos los valores de ϑ, cuando éste exista. Por ejemplo, si tomamos como función de pérdida la fórmula anterior y ϑ es la media de una población normal $N(\vartheta, \sigma)$ con σ conocido, el riesgo del estimador media muestral $\overline{X} = \hat{\vartheta}$ es σ^2/n, y es menor que para cualquier otro estimador, sea cual sea el valor de ϑ.

En la estimación bayesiana al existir siempre una distribución de probabilidad para el parámetro el problema está siempre resuelto. Si llamamos como antes $\ell(\vartheta, \hat{\vartheta})$ a la función de pérdida de oportunidad, el estimador óptimo inicial es aquel que minimiza la pérdida esperada

$$\int \ell(\vartheta, \hat{\vartheta}) f(\vartheta)\, d\vartheta$$

al tomar la muestra $\mathbf{X}$, el estimador óptimo será el que minimice la nueva pérdida esperada:

$$\int \ell(\vartheta, \hat{\vartheta}) f(\vartheta|\mathbf{X})\, d\vartheta$$

donde $f(\vartheta \mid \mathbf{X})$ es la distribución posterior.

El enfoque decisional se adapta mejor a la metodología bayesiana por dos razones: en primer lugar conduce siempre a un estimador claramente definido y óptimo con el criterio elegido; en segundo, establece una guía clara de cómo escoger el estimador, tanto antes como después de tomar la muestra, y de evaluar los beneficios aportados por ésta: con utilidad cuadrática $(\hat{\vartheta} - \vartheta)^2$, la utilidad esperada es la varianza y el valor de la información perfecta es la reducción de varianza entre la posterior y la prior. Dentro del marco clásico el enfoque decisional no supone ninguna ventaja práctica, ya que, en general, no es posible encontrar estimadores con menor riesgo para cualquier valor del parámetro. Esto obliga a cambiar el criterio de decisión o a incluir criterios adicionales (estimadores centrados, invariantes, etc.).

11.7.2 Contrastes y decisiones

Un contraste de hipótesis puede analizarse como un problema de decisión con dos acciones posibles: a_0 = aceptar H_0; a_1 = aceptar H_1. Las consecuencias pueden medirse por una función de pérdida de oportunidad $\ell(a_iH_j)$ tal que $\ell(a_iH_i) = 0$. Entonces, la decisión óptima será a_0 si tiene la menor pérdida esperada, es decir, si:

$$P(H_1)\ell(a_0, H_1) < P(H_0)\ell(a_1, H_0)$$

Como el enfoque clásico no asigna probabilidades a las hipótesis, esta formulación no presenta ventajas especiales. Sin embargo, en el enfoque bayesiano aceptaremos H_0 cuando:

$$\frac{P(H_0)}{P(H_1)} > \frac{\ell(a_0, \mathrm{H}_1)}{\ell(a_1, \mathrm{H}_0)}$$

Supongamos que se toma una muestra N y calculemos las verosimilitudes de obtener el resultado muestral M en función de cada hipótesis. Esto implica que $P(M|H_0)$ y $P(M|H_1)$ son conocidas, y las probabilidades a posteriori de cada hipótesis se obtendrán con el teorema de Bayes de la forma habitual:

$$P(H_i|M) = \frac{P(M|H_i)P(H_i)}{P(M)} \qquad i = 1, 2$$

La estructura del problema de decisión se presenta ahora en la figura 11.4, donde se ha tenido en cuenta que la pérdida, o coste, asociada a la acción correcta es cero. Como antes, la opción a_0 = aceptar H_0 será preferida a la a_1 si:

$$P(H_1|M)\ell(a_0H_1) < P(H_0|M)\ell(a_1H_0)$$

que equivale a:

$$\frac{P(H_0|M)}{P(M|H_1)} = \frac{P(M|H_1)P(H_0)}{P(M|H)P(H_1)} > \frac{\ell(a_0, H_1)}{\ell(a_1, H_0)}$$

es decir:

$$\frac{P(M|H_0)}{P(M|H_1)} > \frac{\ell(a_0, H_1)P(H_1)}{\ell(a_1, H_0)P(H_0)}$$

que puede interpretarse diciendo que tomaremos a_0 (aceptaremos H_0) cuando el cociente de verosimilitudes a su favor sea mayor que el producto de los ratios de las consecuencias de los errores por las probabilidades a priori.

Si suponemos $P(H_1) = P(H_0)$ y las consecuencias de ambos errores idénticas, entonces aceptaremos H_0 —tomaremos la acción a_0— cuando el cociente

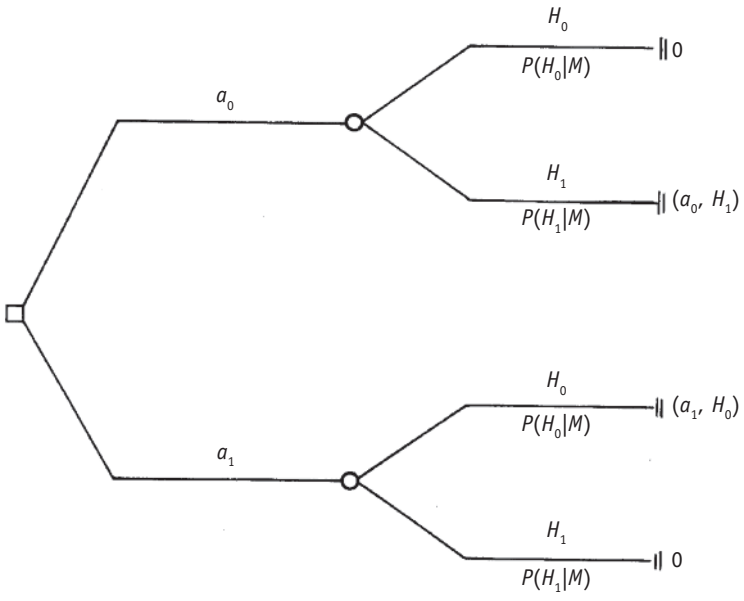
$$\lambda = \frac{P(M|H_0)}{P(M|H_1)} \tag{11.17}$$

sea mayor que uno. En la práctica ambos tipos de error no son iguales, por lo que aceptaremos H_0 cuando:

$$\lambda = \frac{P(M|H_0)}{P(M|H_1)} > k$$

donde k depende de los costes y de las probabilidades a priori. Como (11.17) es el contraste de verosimilitudes, el enfoque bayesiano permite dar una justificación formal a la elección de k.

Figura 11.4 Estructura de un contraste de hipótesis bayesiano



11.8 Resumen del capítulo

Los problemas de decisión en condiciones de incertidumbre admiten una solución general: si aceptamos unos principios generales de coherencia debemos tomar la opción que conduzca a la mayor utilidad esperada. En decisiones repetitivas, y con cantidades que no son importantes para el decisor, la función de utilidad es lineal, y maximizar la utilidad equivale a maximizar el beneficio esperado o a minimizar el coste esperado. En otros casos hay que tener en cuenta las preferencias del decisor. Para analizar problemas complejos el árbol de decisión es una herramienta muy útil, y cuando las probabilidades y las consecuencias sean conocidas, permite obtener la mejor decisión mediante el algoritmo de promediar y retroceder que consiste en sustituir los puntos aleatorios por su esperanza y en cada punto de decisión tomar la estrategia que conduzca al mayor valor esperado.

Los problemas de estimación y contrastes pueden tratarse unificadamente como problemas de decisión en condiciones de incertidumbre. La solución de un problema de decisión es conceptualmente sencilla: escoger aquella alternativa que conduzca a una utilidad esperada mayor. El enfoque de decisión se adapta especialmente bien a la metodología bayesiana de inferencia y permite evaluar el coste de incertidumbre y el valor esperado de la información muestral.

11.9 Lecturas recomendadas

La teoría bayesiana de la decisión se trata en Schlaifer (1969), Lindley (1991), Raiffa (1997), Berger (1993) y Bernardo y Smith (2000), entre otros. El enfoque clásico de la inferencia como un problema de decisión, en Ferguson (1967). La teoría de la utilidad, en Keeney y Raiffa (1993).

12. Diagnosis y crítica del modelo

Karl Pearson (1857-1936)
Científico británico. Inventor del contraste que lleva su nombre y uno de los fundadores de la estadística en el siglo XIX. Fue catedrático de matemáticas y después de eugenesia en la Universidad de Londres. Fundador con Weldon, y con el apoyo económico de Galton, de la prestigiosa revista de estadística *Biometrika.*

12.1 Introducción

Al estimar los parámetros del modelo se supone que los datos constituyen una muestra aleatoria de una distribución que, salvo por sus parámetros, es conocida. La etapa de diagnosis y crítica del modelo consiste en estudiar si estas hipótesis básicas estructurales no están en contradicción con la muestra. En concreto:

1) Si la distribución supuesta es consistente con los datos.
2) Si las observaciones son independientes.
3) Si la muestra es homogénea, es decir, si todas las observaciones provienen de la misma población.

Vamos a analizar la importancia de cada una de estas hipótesis estructurales, cómo contrastarlas con los datos y cómo modificar los procedimientos estudiados cuando resulten falsas.

12.2 La hipótesis sobre la distribución

12.2.1 Efecto de un modelo distinto del supuesto

La elección del estimador de los parámetros y la estimación de su precisión, que son los ingredientes básicos para construir intervalos y contrastar hipótesis, dependen del modelo supuesto. Diremos que un procedimiento estadístico es *robusto* frente a una hipótesis cuando es aproximadamente válido ante pequeñas desviaciones de la hipótesis.

Las inferencias respecto a medias son en general robustas: sea cual sea la población base la media muestral es centrada con varianza σ^2/n y, por el teorema central de límite, su distribución es asintóticamente normal. En consecuencia, intervalos de confianza y contrastes basados en la distribución t de Student son aproximadamente válidos con independencia de la distribución de partida, es decir, un intervalo del 95% contendrá, a largo plazo, el 95% de las veces la media de la población.

Sin embargo, si la distribución es falsa, aunque las inferencias respecto a las medias sean válidas, dejan de ser óptimas. Por ejemplo, mostramos en el capítulo 7 (sección 7.5) cómo una pequeña contaminación en una distribución normal hace bajar drásticamente la eficiencia de la media muestral. Por otro lado, la mejor estimación de σ^2 deja de ser $\hat{s}^2$, con lo que no utilizamos adecuadamente la información disponible. El resultado es que los procedimientos que suponen normalidad son, aunque válidos, poco precisos cuando esta hipótesis no es cierta, lo que se traduce en intervalos innecesariamente grandes o contrastes poco potentes. Por ejemplo, si los datos son uniformes (o, b), un intervalo para la media basado en $(x_{\max}/2)$, el estimador *MV,* es en promedio más corto que el basado en $\bar{x}$. Análogamente, si la distribución tiene mucho apuntamiento —como ocurre si mezclamos dos normales con misma media y varianzas muy distintas— la mediana proporcionará intervalos más cortos que la media. Finalmente, con distribuciones asimétricas, la estimación de σ con $\hat{s}$ puede ser muy ineficiente, ya que $\bar{x}$ y $\hat{s}$ son independientes únicamente en la distribución normal y, en general, la estimación óptima de σ requiere utilizar la información de $\bar{x}$. (Por ejemplo, la estimación *MV* de la varianza es para Poisson$\sqrt{\bar{x}}$, y para la exponencial $1/\bar{x}$).

Las inferencias respecto a varianzas son muy sensibles a la hipótesis de normalidad. Para cualquier población $\hat{s}^2$ es un estimador centrado de σ^2, pero su varianza depende mucho del apuntamiento de la distribución base (véase cuadro pág. 280). También su distribución es muy dependiente de la población. En consecuencia, los intervalos o contrastes estudiados para varianzas serán poco precisos si la población no es aproximadamente normal.

La forma de comprobar si los datos provienen de una distribución es efectuar un contraste de ajuste. Los dos contrastes básicos son el χ^2 de Pearson y el Kolmogorov-Smirnov, que presentamos a continuación.

12.2.2 El contraste χ^2 de Pearson

El contraste de ajuste más antiguo es el contraste de Pearson, cuya idea es comparar las frecuencias observadas en un histograma o un diagrama de barras con las especificadas por el modelo teórico que se contrasta. Este contraste es válido para todo tipo de distribuciones, discretas y continuas.

La hipótesis H_0 es que unos datos de una variable x provienen de un determinado modelo. Existen dos variantes posibles. En la primera H_0 especifica completamente la distribución (por ejemplo, x es $N[5, 2]$), en la segunda H_0 especifica la forma, pero no los parámetros, que se estiman a partir de los datos (por ejemplo, x es normal).

Aplicación del test

Supondremos que se dispone de una muestra $\mathbf{X} = (x_1, ..., x_n)$ aleatoria simple de una variable continua o discreta, donde $n \geq 25$. El contraste se realiza como sigue:

1) Agrupar los n datos en k clases, donde $k \geq 5$. Las clases se eligen de manera que cubran todo el rango posible de valores de la variable y que cualquier posible dato quede clasificado sin ambigüedad. Normalmente esto exigirá que los intervalos extremos sean abiertos. Por razones que veremos después, es conveniente tener, aproximadamente, el mismo número de datos en cada clase, y al menos tres datos en cada una. Llamaremos O_i a la *frecuencia observada* en la muestra de la clase i, es decir, el número de datos muestrales en dicha clase.
2) Calcular la probabilidad p_i que el modelo supuesto asigna a cada clase. Como éstas cubren todo el rango de la variable, $\Sigma p_i = 1$, llamaremos:

$$E_i = np_i$$

a la *frecuencia esperada* de la clase i de acuerdo con el modelo.
3) Calcular la discrepancia entre las frecuencias observadas y las previstas por el modelo mediante:

$$X^2 = \sum_{i=1}^{k} \frac{(\text{Observadas}_i - \text{Esperadas}_i)^2}{\text{Esperadas}_i}$$

que se distribuye aproximadamente como una χ^2 cuando el modelo es correcto. Sus grados de libertad son:

a) Si el modelo especifica completamente las probabilidades p_i que son conocidas *antes* de tomar la muestra, el número de grados de libertad será $k - 1$.
b) Si las probabilidades p_i se han calculado estimando r parámetros del modelo por máxima verosimilitud, el número de grados de libertad es $k - r - 1$.

Rechazaremos el modelo cuando la probabilidad de obtener una discrepancia mayor o igual que la observada sea suficientemente baja. Es decir, cuando:

$$X^2 \geq \chi^2_\alpha(k - r - 1)$$

para un cierto α pequeño.

El test no contrasta un modelo concreto, sino la clase de modelos que atribuyen probabilidades iguales al supuesto a los intervalos construidos (figura 12.1). Por esta razón es recomendable que el número de clases sea grande (siempre mayor que cinco).

Un inconveniente del contraste es que al tomar las diferencias $(O_i - E_i)$ al cuadrado es insensible a pautas de variación sistemáticas. Por ejemplo, cuando la distribución supuesta esté situada con relación a la real, como indica la figura 12.2, la secuencia de signos de las diferencias será + + + + + ..., indicando claramente una pauta.

Figura 12.1 Los modelos A y B serán indistinguibles en un contraste χ^2 con tres clases

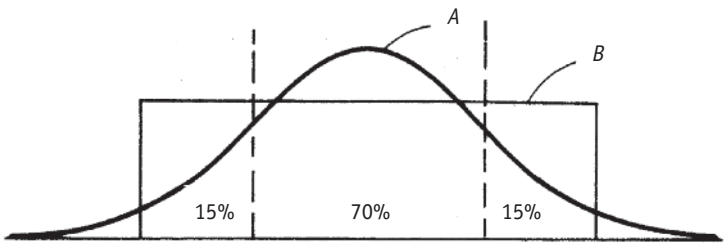
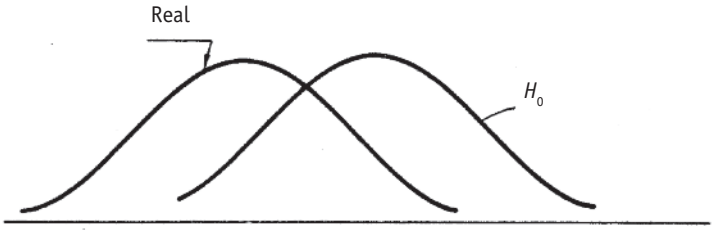


Figura 12.2 Pautas de signos de las diferencias ($O_i - E_i$) en el contraste χ^2



Por último, conviene calcular los términos

$$\frac{(O_i - E_i)^2}{E_i}$$

separadamente para estudiar la contribución de cada clase al rechazo de H_0. Esto permitirá comprender si el modelo se ha rechazado por un único valor aislado o por el conjunto, ya que un solo valor extremo, debido quizás a un error en los datos, puede tener un efecto excesivo en el contraste.

En resumen, si el contraste conduce a rechazar H_0, conviene investigar siempre la causa, para sugerir un modelo alternativo.

Justificación

El lector puede encontrar en el apéndice 12A un análisis más detallado de las propiedades matemáticas del contraste; aquí presentaremos una justificación intuitiva y no rigurosa del mismo.

Sea I_i un intervalo cualquiera de valores de la variable, y clasifiquemos los n datos muestrales en dos clases: dentro o fuera del intervalo I_i. Sea O_i la variable aleatoria que cuenta el número de elementos dentro de I_i (su frecuencia observada). Al tomar muchas muestras, O_i tendrá una distribución binomial, con esperanza $np_i = E_i$ y desviación típica $\sqrt{np_iq_i}$. Cuando n es grande suponiendo p_i pequeño, O_i será, aproximadamente, Poisson, con $\lambda = np_i$; si $\lambda > 5$ utilizaremos la aproximación normal:

$$\left(\frac{O_i - np_i}{\sqrt{np_i}}\right)^2 \sim N(0, 1)^2$$

Las variables O_i son dependientes, ya que están unidas por la restricción lineal $\Sigma O_i = n$. Por tanto, solamente son independientes $k - 1$ y al sumar sus cuadrados se obtiene una χ^2 con $k - 1$ grados de libertad.

Cuando estimamos r parámetros directamente a partir de las frecuencias O_i, para calcular las p_i, establecemos r restricciones adicionales sobre las O_i. Por ejemplo, si estimamos la media:

$$\bar{x} = \frac{\Sigma x_i O_i}{n} = \Sigma c_i O_i$$

que supone una restricción lineal adicional. Por tanto, tendremos únicamente $k - 1 - r$ variables independientes, que serán los grados de libertad de la χ^2.

Cuando la estimación de los parámetros utiliza las clases en lugar de los datos originales —como ocurre con variables continuas—, cada parámetro no impone ya necesariamente una restricción sobre los O_i, y el problema se complica. Una solución aproximada es tomar entonces $k - r - 1$ grados de libertad, suponiendo que estimamos los parámetros por máxima verosimilitud.

El razonamiento anterior exige que $np_i = E_i$ sea mayor que 5, para que la aproximación normal sea razonable. Además el tamaño muestral debe ser como mínimo 30.

Ejemplo 12.1

Durante la Segunda Guerra Mundial se dividió el mapa de Londres en cuadrículas de 1/4 km² y se contó el número de bombas caídas en cada cuadrícula durante un bombardeo alemán. Los resultados fueron:

x_i: Impactos en la cuadrícula	0	1	2	3	4	5
O_i: Frecuencia	229	211	93	35	7	1

Contrastar la hipótesis de que los datos siguen una distribución de Poisson.

Solución:

El valor estimado de λ es:

$$\lambda = \frac{\Sigma x_i O_i}{\Sigma O_i} = \frac{535}{576} = 0{,}929$$

Como:

$$p_i = e^{-0,929} \cdot (0,929)^i/i! \qquad i = 0, 1, \ldots, 5$$

entonces las frecuencias esperadas son:

$$E_0 = 0,395 \cdot 576 = 227,5;\ E_1 = 211;\ E_2 = 98;\ E_3 = 30;\ E_4 = 7;\ E_5 = 1,5$$

El estadístico es:

$$X^2 = \frac{(229 - 227,5)^2}{227,5} + \ldots + \frac{(1 - 1,5)^2}{1,5} =$$

$$= 0,01 + 0 + 0,26 + 0,83 + 0 + 0,17 = 1,27$$

Si la distribución de Poisson es adecuada, X^2 es un valor de una χ^2 con $6 - 2 = 4$ grados de libertad y no hay razón para dudar de la hipótesis. El ajuste es muy bueno.

El hecho de que los datos sigan una distribución de Poisson sugiere que el bombardeo era aleatorio y no dirigido a determinados objetivos militares.

Ejemplo 12.2

La vida de 70 motores ha tenido la siguiente distribución:

Años de funcionamiento	(0,1)	(1,2)	(2,3)	(3,4)	Más de 4
Frecuencia	30	23	6	5	6

donde el intervalo (a, b) indica una vida t: $a \leq t < b$. ¿Puede suponerse que su vida sigue la distribución exponencial? Para calcular la media supondremos que el intervalo de más de 4 tiene el centro en 5. Entonces:

$$\bar{x} = \frac{\Sigma x_i o_i}{\Sigma o_i} = 0,5\,\frac{30}{70} + 1,5\,\frac{23}{70} + \frac{6}{70} + 3,5\,\frac{5}{70} + 5\,\frac{6}{70} = 1,60$$

Para calcular probabilidades utilizaremos la función de distribución:

$$F(x) = 1 - e^{-\frac{x}{1,60}}$$

$F(1) = 0{,}46$; $F(2) = 0{,}71$; $F(3) = 0{,}84$; $F(4) = 0{,}92$.
$E_1 = 70 \cdot 0{,}46 = 32{,}2$; $E_2 = 70(0{,}71 - 0{,}46) = 17{,}5$; $E_3 = 70(0{,}84 - 0{,}71) = 9{,}10$;
$E_4 = 70 \cdot (0{,}92 - 0{,}84) = 5{,}60$; $E_5 = 70 \cdot 0{,}08 = 5{,}6$.

Por tanto:

$$X^2 = \frac{(30 - 32{,}2)^2}{32{,}2} + \ldots + \frac{(6 - 5{,}6)^2}{5{,}6} = 3{,}03$$

Como x2(3; 0,05) = 7,81 y x2 (4; 0,05) = 9,48 no hay evidencia para rechazar la distribución exponencial.

12.2.3 El contraste de Kolmogorov'''Smirnov

Este contraste compara la función de distribución teórica con la empírica. Es válido únicamente para variables continuas.

Aplicación del contraste

La hipótesis nula en este contraste es que la muestra proviene de un modelo continuo $F(x)$. El procedimiento para construir el contraste es:

1) Ordenar los valores muestrales, de manera que:

$$x_{(1)} \leq x_{(2)} \ldots \leq x_{(n)}$$

2) Calcular la función de distribución empírica de la muestra, $F_n(x)$, con:

$$F_n(x) = \begin{cases} 0, & \text{si} \quad x < x_{(1)} \\ \dfrac{r}{n}, & \text{si} \quad x_{(r)} \leq x < x_{(r+1)} \\ 1, & \text{si} \quad x \geq x_{(n)} \end{cases}$$

3) Calcular la discrepancia máxima entre las funciones de distribución observada (o empírica) y teórica con el estadístico:

$$D_n = \text{máx}|F_n(x) - F(x)|$$

cuya distribución, cuando $F(x)$ es cierta, se ha tabulado (véase la tabla 8). Si la distancia calculada D_n es mayor que la encontrada en las tablas, fijado α, rechazaremos el modelo $F(x)$.

Este contraste tiene la ventaja de que no requiere agrupar los datos y el inconveniente de que si calculamos $F(x)$ estimando parámetros de la población mediante la muestra, la distribución de D_n es sólo aproximada: el contraste es conservador, tendiendo a aceptar H_0. En contrapartida, permite construir bandas de confianza de la distribución: si $D(\alpha, n)$ es el valor obtenido en tablas para D_n, tendremos que, si F es correcta, con confianza $1 - \alpha$:

$$D(\alpha, n) \geq \text{máx}|F_n(x) - F(x)|$$

por tanto:

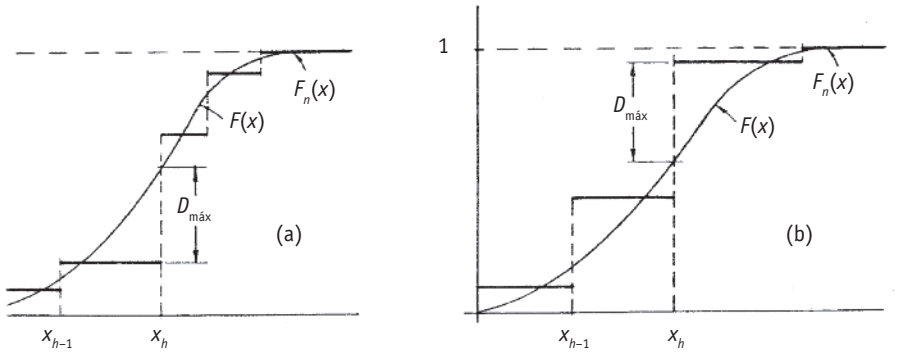
$$F(x) \in [F_n(x) \pm D(\alpha, n)]$$

y llevando $D(\alpha, n)$ a ambos lados de $F_n(x)$ se obtienen bandas de confianza para la distribución.

La figura 12.3 muestra los dos casos que pueden presentarse:

a) La distancia máxima entre $F(x)$ y $F_n(x)$ se da inmediatamente antes de llegar a x_h y su magnitud es $|F_n(x_{h-1}) - F(x_h)|$.
b) La distancia máxima es $|F_n(x_h) - F(x_h)|$.

Figura 12.3 Aplicación del contraste de Kolmogorov-Smirnov



Por tanto, al aplicar el test hay que calcular para cada punto x_h:

$$D_n(x_h) = \text{máx}\{|F_n(x_{h-1}) - F(x_h)|, |F_n(x_h) - F(x_h)|\}$$

y tomar el máximo después de estos $D_n(x_h)$

Justificación del test

Este test se basa en que la distribución de D_n es la misma sea cual sea la distribución de partida. Para demostrarlo, supongamos que se desea contrastar que los datos provienen de una población $F(x)$ continua completamente especificada (por ejemplo, $N[10, 2]$) que supondremos para simplificar es monótona creciente [si $x_1 < x_2$, $F(x_1) < F(x_2)$]. Construyamos una nueva variable mediante la transformación:

$$y = F(x) \tag{12.1}$$

que según la sección 5.7.2 se distribuye como una uniforme en el intervalo (0, 1), sea cual sea la distribución *F.* Con esta transformación, la muestra $\mathbf{X} = (x_1, \ldots. x_n)$ se convierte en una muestra aleatoria simple de una uniforme (0, 1). Si x_j es el elemento *j*-ésimo con $F_n(x_j) = j/n$, su transformado:

$$y_j = F(x_j)$$

será también el elemento *j*-ésimo con función de distribución empírica $G_n(y_j) = j/n$, y, para todos los datos:

$$G_n(y_j) = F_n(x_j) = j/n$$

Entonces, si los datos **X** siguen la distribución *F,* los datos **Y** generados con (12.1) seguirán la distribución $U(0, 1)$ con función de distribución $G(y) = y$. Por tanto, si la hipótesis que contrastamos es cierta:

$$|G(y) - G_n(y)| = |y - G_n(y)| = |F(x) - F_n(x)|$$

de manera que:

$$D_n = \max |F(x) - F_n(x)| = \max |y - G_n(y)|$$

y para obtener la distribución de D_n basta estudiar la distancia entre la recta $G(y) = y$, y la distribución empírica en muestras aleatorias de tamaño *n* de una variable uniforme (0, 1). Esta distribución se obtiene fácilmente con el método de Montecarlo y está tabulada en la tabla 8 del apéndice de tablas.

Ejemplo 12.3

Contrastar si la muestra siguiente de duraciones de vida puede suponerse exponencial:

16, 8, 10, 12, 6, 10, 20, 7, 2, 24

La media es $\bar{x} = 11{,}50$. Por tanto:

$$F(x) = 1 - e^{-\frac{x}{11{,}5}}$$

Construiremos la tabla

x	$F_n(x)$	$F(x)$	$D_n(x)$
2	0,1	0,16	0,16
6	0,2	0,41	0,31
7	0,3	0,46	0,26
8	0,4	0,50	0,20
10	0,6	0,58	0,18
12	0,7	0,65	0,05
16	0,8	0,75	0,05
20	0,9	0,82	0,08
24	1	0,88	0,12

El valor máximo de D_n es 0,31. En la tabla 8 se obtiene con $n = 10$: $D(0{,}2;\ 10) = 0{,}323$; $D(0{,}1;\ 10) = 0{,}369$, con lo que el nivel crítico p del contraste es aproximadamente 0,2 y aceptaremos la distribución exponencial.

12.2.4 Contrastes de normalidad

Por su importancia vamos a estudiar con detalle el problema de contrastar la normalidad. Además de los dos estudiados, existen otros tres tipos de contrastes:

1) Mediante el ajuste del diagrama probabilístico-normal a una recta. Éste es el contraste de Shapiro y Wilk.

2) Por las medidas de asimetría y apuntamiento de los datos.
3) Estimando la transformación necesaria para conseguir normalidad.

No existe un contraste «óptimo» para probar la hipótesis de normalidad. La razón es que la potencia relativa depende del tamaño muestral y de la verdadera distribución que genera los datos. Desde un punto de vista poco riguroso, el contraste de Shapiro y Wilks es, en términos generales, el más conveniente en pequeñas muestras ($n < 30$), mientras que el contraste χ^2 de Pearson y el de Kolmogorov-Smirnov, en la versión modificada en Lilliefors (1967), son adecuados para muestras grandes.

Cuando se sospeche que hay desviaciones de la normalidad en una dirección conocida pueden utilizarse los contrastes de asimetría y curtosis. Finalmente el contraste sobre la transformación es adecuado cuando se pretenda transformar, como veremos en la sección 12.2.6.

El contraste de Shapiro y Wilk

Este contraste mide el ajuste de la muestra representada en papel probabilístico normal a una recta. Se rechaza la normalidad cuando el ajuste es malo, que corresponde a valores pequeños del estadístico. La justificación del contraste se presenta en el apéndice 12B. El estadístico es:

$$w = \frac{1}{ns^2}\left[\sum_{j=1}^{h} a_{j,n}(x_{[n-j+1]} - x_{[j]})\right]^2 = \frac{A^2}{ns^2}$$

donde $ns^2 = \Sigma(x_i - \bar{x})^2$; h es $n/2$ si n es par y $(n-1)/2$ si n es impar; los coeficientes $a_{j,n}$ están tabulados en el apéndice (tabla 10) y $x_{(j)}$ es el valor ordenado en la muestra que ocupa el lugar j. La distribución de w está tabulada (tabla 11) y se rechaza la normalidad cuando el valor calculado es *menor* que el valor crítico dado en las tablas. La razón es que w mide el ajuste a la recta, y no la discrepancia con la hipótesis.

Ejemplo 12.4

Contrastar la hipótesis de que los datos siguientes provienen de una distribución normal: (20, 22, 24, 30, 31, 32, 38).

Para aplicar el test calcularemos los valores $a_{i,n}$ directamente de la tabla 10 del apéndice; entonces:

$$a_{17} = 0{,}6233$$
$$a_{27} = 0{,}3031$$
$$a_{37} = 0{,}1401$$

Por lo tanto, A será:

$$A = a_{17}[x_{(7)} - x_{(1)}] + a_{27}[x_{(6)} - x_{(2)}] + a_{37}[x_{(5)} - x_{(3)}] =$$

$$= 0{,}6233 \cdot (18) + 0{,}3031(10) + 0{,}1401(7) = 15{,}2311.$$

Como

$$s^2 = 34{,}9796, \qquad ns^2 = 244{,}8571$$

$$A^2 = 231{,}9864$$

el estadístico resultante será:

$$\omega = \frac{231{,}9864}{244{,}8571} = 0{,}9474$$

El valor de ω para $n = 7$ y un nivel de significación de 0,05 es 0,803, menor que el obtenido, por lo que aceptamos la hipótesis de normalidad.

El contraste de Kolmogorov-Smirnov-Lilliefors

El contraste *KS* utiliza el estadístico:

$$D_n = \sup|F_n(x) - F(x)|$$

donde $F_n(x)$ es la función de distribución empírica muestral y $F(x)$ la teórica de la población que queremos contrastar. Ha sido extensamente tabulado en el supuesto de que la distribución $F(x)$ queda totalmente especificada con la hipótesis. Ya comentamos que cuando se estiman parámetros para calcular $F(x)$, la tabulación clásica de este test conduce a un contraste muy conservador, siendo el nivel de significación mucho más bajo que el dado por la tabla. Lilliefors ha tabulado este estadístico cuando estimamos los parámetros μ y σ^2 de la distribución normal con $\bar{x}$ y $\hat{s}^2$ (véase tabla 9).

El contraste se efectúa calculando el estadístico D y rechazando la hipótesis de normalidad cuando el valor de D obtenido es significativamente grande, es decir, mayor que el valor dado por las tablas al nivel de significación escogido.

La potencia de este contraste para tamaños muestrales medianos es baja. Por ejemplo, para detectar la diferencia entre una $N(0,1)$ y una uniforme en $[-\sqrt{3}, \sqrt{3}]$ se necesitan más de 100 datos.

Ejemplo 12.5

Aplicaremos el contraste a los datos del ejemplo anterior. En primer lugar calcularemos la media y varianza muestrales para obtener las probabilidades teóricas.

La media de los siete datos es:

$$\bar{x} = 28{,}14$$

y la desviación típica, corregida por grados de libertad:

$$\hat{s} = \frac{\Sigma(x_i - \bar{x})^2}{n - 1} = 6{,}39$$

Para efectuar el test construimos la tabla:

N.º de Obs.	x	$F_n(x)$	$F(x)$	D_1	D_2	D_n
1	20	0,1429	0,1020	0,1020	0,0409	0,1020
2	22	0,2857	0,1685	0,0256	0,1172	0,1172
3	24	0,4286	0,2578	0,0279	0,1718	0,1718
4	30	0,5714	0,6141	0,1855	0,0427	0,1855
5	31	0,7143	0,6736	0,1018	0,0407	0,1018
6	32	0,8571	0,7258	0,1115	0,0313	0,1115
7	38	1,0000	0,9382	0,0811	0,0618	0,0811

Los valores de $F_n(x)$ se han obtenido simplemente mediante

$$F_n(x_i) = \frac{i}{n}$$

donde i representa el índice ordinal de la observación, y los valores $F(x)$ se han calculado tipificando los siete datos y mirando en tablas de la normal estándar. Para ello se calcula:

$$z_i = \frac{x_i - \bar{x}}{\hat{s}}$$

y se obtiene en tablas el valor de $F(z_i)$. En la tabla, D_1 representa el valor:

$$D_1 = |F_n(x_{i-1}) - F(x_i)|$$

mientras que D_2 es:

$$D_2 = |F_n(x_i) - F(x_i)|$$

La tabla de Lilliefors del anexo indica que el valor de D crítico para un nivel de significación del 5% y un tamaño muestral de 7 es:

$$D_c = (0{,}05;\ 7) = 0{,}300$$

y como la máxima distancia obtenida en nuestros datos es:

$$D_n = 0{,}1855$$

concluimos que no hay evidencia suficiente en los datos para rechazar la hipótesis de normalidad.

Contraste χ^2 de Pearson (contraste *CS*)

El contraste de normalidad mediante el estadístico χ^2 de Pearson estudiado en 12.2.2 se efectúa siguiendo las reglas allí expuestas. Su aplicación más frecuente es a problemas donde μ y σ^2 se estiman a partir de los datos, con lo que la distribución resultante tendrá, aproximadamente, $k - 3$ grados de libertad, siendo k el número de clases.

El contraste no especifica cómo seleccionar las clases, aspecto que ha sido objeto de una abundante literatura. La regla más extendida es tomar clases equiprobables, en número tal que la frecuencia teórica de cada clase sea mayor que 3. El contraste χ^2 de Pearson suele utilizarse únicamente en el caso de muestras grandes.

Para facilitar la realización del test reproducimos en la tabla 12.1 los cuantiles de la distribución normal estándar que son necesarios para efectuar el contraste con clases equiprobables. Con 8 clases, cada una de ellas debe contener una probabilidad de (1/8), y la tabla proporciona directamente los valores para construir 8, 6, 5 y 4 clases equiprobables. Por ejemplo, para hacer ocho clases equiprobables tomaremos como límites:

$$(-\infty;\ \bar{x} - 1{,}15s);\ (\bar{x} - 1{,}15s;\ \bar{x} - 0{,}68s);\ (\bar{x} - 0{,}68s;\ \bar{x} - 0{,}32s);$$

$$(\bar{x} - 0{,}32s; \bar{x}); (\bar{x}; \bar{x} + 0{,}32s); (\bar{x} + 0{,}32s; \bar{x} + 0{,}68s);$$

$$(\bar{x} + 0{,}68s; \bar{x} + 1{,}15s); (\bar{x} + 1{,}15s; +\infty)$$

Con k clases equiprobables la frecuencia teórica de cada clase es n/k y la fórmula de X^2 se reduce a:

$$X^2 = \frac{k}{n} \Sigma O_i^2 - n$$

Conviene aplicar este contraste cuando n sea ≥ 100. Para tamaños muestrales menores el contraste no rechazará la normalidad para casi cualquier distribución simétrica con un único máximo.

Tabla 12.1 Cuantiles en la distribución normal para el test χ^2

$P(x \leq x_p)$	1/8	1/6	1/5	1/4	1/3	3/8	2/5
x_p	–1,15	–0,97	–0,84	–0,68	–0,43	–0,32	–0,26

Ejemplo 12.6

Comprobar si los datos siguientes provienen de una distribución normal: 107,9; 96,7; 91,2; 79; 103,1; 88; 101,3; 106; 93,7; 86; 100,7; 99,4; 104,6; 117,2; 112,2; 106,9; 93; 88,3; 101,9; 109,8.

Aunque teniendo veinte datos el test de Shapiro y Wilk será, en general, preferido, utilizaremos el contraste χ^2 como ejemplo.

En primer lugar calcularemos la media y desviación típica de los datos:

$$\bar{x} = \frac{\Sigma x_i}{n} = 99{,}35$$

$$s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}} = 9{,}56$$

Por lo tanto, si decidimos hacer cinco clases equiprobables, los límites inferiores de cada clase serán:

$$\bar{x} - 0{,}84s = 91{,}32$$
$$\bar{x} - 0{,}26s = 96{,}87$$
$$\bar{x} + 0{,}26s = 101{,}83$$
$$\bar{x} + 0{,}84s = 107{,}38$$

Con lo que obtendríamos la clasificación:

Clase		O_i	E_i	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
$-\infty$	91,32	5	4	1	0,25
91,32	96,87	3	4	1	0,25
96,87	101,83	3	4	1	0,25
101,83	107,36	5	4	1	0,25
107,36	∞	4	4	0	0
TOTAL		20	20		1

$$X^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} = 1$$

Como el valor crítico de una χ^2 con 2 grados de libertad al nivel de significación $\alpha = 0{,}05$ es:

$$X^2_\alpha = 5{,}99$$

aceptamos la hipótesis de normalidad. (Los datos del ejemplo se han generado de una distribución normal con media 100 y desviación típica 10.)

Contrastes de asimetría y curtosis

El coeficiente de asimetría muestral que definimos en la sección 2.4:

$$CA = \alpha_1 = \frac{\Sigma(x_i - \bar{x})^3}{ns^3}$$

estima un parámetro de la población que es cero si la hipótesis de normalidad es cierta. Para muestras grandes —como mínimo 50 datos— la distribución de α_1 es aproximadamente normal con media y varianza:

$$\mathrm{E}(\alpha_1) = 0 \qquad \mathrm{Var}(\alpha_1) \simeq \frac{6}{n}$$

lo que nos permite contrastar la hipótesis de que los datos provienen de una distribución simétrica.

El grado de apuntamiento o curtosis —concentración de probabilidad en el centro frente a las colas— se mide por el coeficiente:

$$CAp = \alpha_2 = \frac{\Sigma(x_i - \bar{x})^4}{ns^4}$$

y toma el valor 3 para una distribución normal. Para muestras grandes —más de 200 observaciones—, α_2 se distribuye asintóticamente normal con media 3 (valor teórico del coeficiente de curtosis en una distribución normal) y varianza:

$$\mathrm{Var}(\alpha_2) \simeq \frac{24}{n}$$

Podemos combinar ambas medidas en un contraste conjunto construyendo el estadístico

$$\frac{n\alpha_1^2}{6} + \frac{n(\alpha_2 - 3)^2}{24} = X_2^2$$

que se distribuye asintóticamente como una χ^2 con dos grados de libertad.

12.2.5 Soluciones

Supongamos que la hipótesis de normalidad es rechazada por los datos, o que, sin serlo totalmente, hay cierta evidencia de que puede no ser cierta. La solución a adoptar depende del tipo de distribución que muestran los datos:

a) si la distribución es unimodal y asimétrica, la solución más simple y efectiva suele ser transformarlos para convertirlos en normales;
b) si la distribución es más apuntada que la normal, o muestra valores atípicos, investigar la presencia de heterogeneidad en los datos (sección 12.4). Como solución global utilizar estimadores robustos;
c) si la distribución es bimodal, investigar la presencia de heterogeneidad. En este caso ni la transformación ni los métodos robustos serán

de mucha utilidad si no segmentamos antes la población en subpoblaciones homogéneas;

d) cuando el objetivo no sea estimar los parámetros sino conocer la distribución, pueden utilizarse métodos no paramétricos para estimar la densidad, como veremos a continuación.

Vamos a analizar las soluciones (a) y (d) con cierto detalle.

12.2.6 Transformaciones para conseguir la normalidad

Box y Cox (1964) han sugerido la siguiente familia de transformaciones para conseguir la normalidad:

$$x^{(\lambda)} = \begin{cases} \dfrac{(x+m)^{\lambda}-1}{\lambda} & (\lambda \neq 0) \qquad (x > -m) \\ \ln (x+m) & (\lambda = 0) \qquad (m > 0) \end{cases}$$

donde λ es el parámetro de la transformación que se estima a partir de los datos y la constante m se elige de forma que $x + m$ sea siempre positiva. Por lo tanto, m será cero si trabajamos con datos positivos e igual en valor absoluto al valor más negativo observado, en otro caso.

Suponiendo $m = 0$ la figura 12.4 presenta la familia de transformaciones para varios valores de λ. Es fácil ver que el logaritmo es la transformación límite cuando λ tiende a cero. Escribiendo x^{λ} como $e^{\lambda \ln x}$, tendremos:

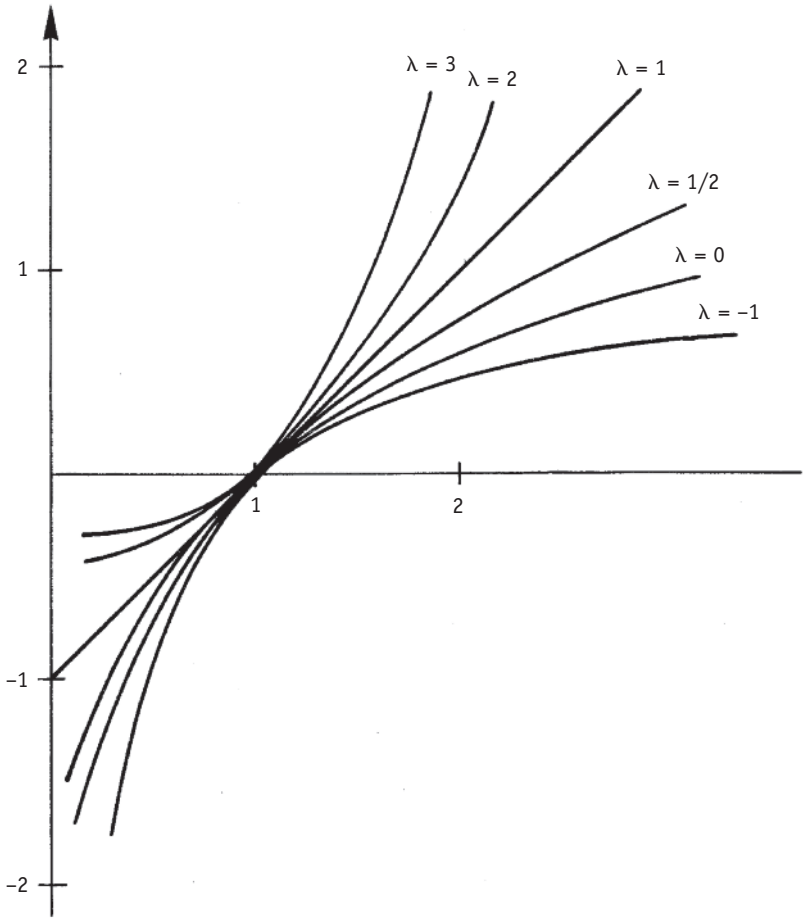
$$x^{(\lambda)} = \frac{e^{\lambda \ln x} - 1}{\lambda}$$

y, cuando λ tiende a cero, la transformación queda indeterminada. Utilizando la regla de L'Hospital y derivando numerador y denominador:

$$\lim_{\lambda \to 0} x^{(\lambda)} = \lim_{\lambda \to 0} \frac{e^{\lambda \ln x} \cdot \ln x}{1} = \ln x$$

Por tanto, esta familia incluye como casos particulares la transformación logarítmica, la raíz cuadrada y la inversa. Se observa en la figura 12.4 que cuando $\lambda > 1$, la transformación produce una mayor separación o dispersión de los valores grandes de x, tanto más acusada cuanto mayor sea el valor de λ, mientras que cuando $\lambda < 1$ el efecto es el contrario: los valores de x grandes tienden a concentrarse, y los valores pequeños ($x < 1$), a dispersarse. Los aspectos básicos de estas transformaciones se presentaron en el

Figura 12.4 Representación gráfica de la familia Box-Cox con $m = 0$ y distintos valores de λ



capítulo 2. Vamos a estudiar en esta sección la estimación del parámetro λ a partir de la muestra. En el apéndice 12C se introduce un método gráfico que puede ser útil si no se dispone de medios de cálculo adecuados.

La estimación *MV* de la transformación

Supongamos que $m = 0$ y que existe un valor de λ que transforma a la variable en normal. La relación entre el modelo para los datos originales x y para los transformados $x^{(\lambda)}$ será:

$$f(x) = f[x^{(\lambda)}] \left| \frac{dx^{(\lambda)}}{dx} \right|$$

como:

$$\frac{dx^{(\lambda)}}{dx} = \frac{\lambda x^{\lambda-1}}{\lambda} = x^{\lambda-1}$$

y suponiendo que $x^{(\lambda)}$ es $N(\mu, \sigma)$ para cierto λ, la función de densidad de las variables originales será:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\left(\frac{x^\lambda - 1}{\lambda} - \mu\right)^2} \cdot x^{\lambda-1}$$

Por tanto, la función de densidad conjunta de $\mathbf{X} = (x_1, ..., x_n)$ será, por la independencia de las observaciones:

$$f(\mathbf{X}) = \frac{1}{\sigma^n(\sqrt{2\pi})^n} \left(\prod_{i=1}^{n} x_i^{\lambda-1}\right) e^{-\frac{1}{2\sigma^2}\Sigma\left(\frac{x_i^\lambda - 1}{\lambda} - \mu\right)^2}$$

y la función soporte o logaritmo de la verosimilitud

$$L(\lambda; \mu, \sigma^2) = -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln 2\pi + (\lambda - 1)\,\Sigma \ln x_i - \frac{1}{2\sigma^2}\Sigma\left(\frac{x_i^\lambda - 1}{\lambda} - \mu\right)^2$$

Para obtener el máximo de esta función utilizaremos que, para λ fijo, los valores de σ^2 y μ que maximizan la verosimilitud (o el soporte) son, derivando e igualando a cero:

$$\hat{\sigma}^2(\lambda) = \frac{1}{n}\Sigma\,[x^{(\lambda)} - \hat{\mu}(\lambda)]^2$$

$$\hat{\mu}(\lambda) = \bar{x}^{(\lambda)} = \Sigma\,\frac{x_i^{(\lambda)}}{n} = \frac{1}{n}\Sigma\left(\frac{x_i^\lambda - 1}{\lambda}\right)$$

Al sustituir estos valores en la verosimilitud obtenemos lo que se denomina la función de verosimilitud *concentrada* en λ. Su expresión es, prescindiendo de constantes:

$$L(\lambda; \hat{\mu}, \hat{\sigma}^2) = -\frac{n}{2}\ln\hat{\sigma}(\lambda)^2 + (\lambda - 1)\,\Sigma \ln x_i \qquad (12.2)$$

Se obtiene una expresión más simple llamando $\dot{x}$ a la media geométrica de las observaciones:

$$\ln \dot{x} = \frac{1}{n}\Sigma \ln x_i$$

con lo que la expresión anterior puede escribirse:

$$L(\lambda) = -\frac{n}{2} \ln [\hat{\sigma}(\lambda)/\dot{x}^{\lambda-1}]^2 = -\frac{n}{2} \ln \left[\frac{1}{n} \Sigma \left(\frac{x^{(\lambda)}}{\dot{x}^{\lambda-1}} - \frac{\hat{\mu}^{(\lambda)}}{\dot{x}^{\lambda-1}} \right)^2 \right]$$

y definiendo la variable

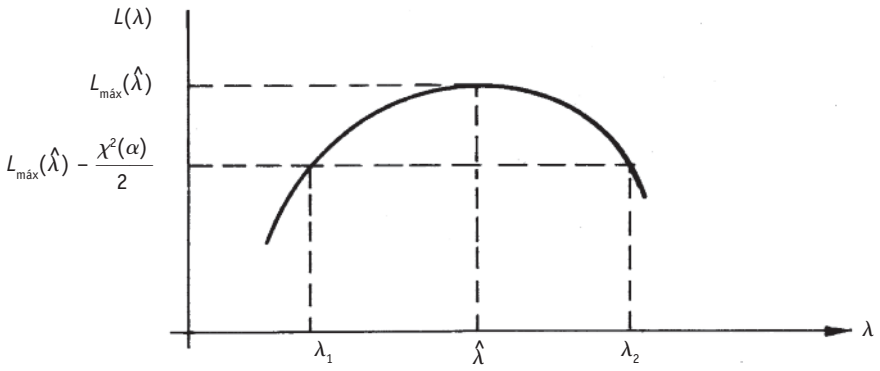
$$z(\lambda) = \frac{x^{\lambda} - 1}{\lambda \dot{x}^{\lambda-1}}$$

concluimos que:

$$\boxed{L(\lambda) = -\frac{n}{2} \ln \left(\frac{1}{n} \Sigma [z_i(\lambda) - \bar{z}(\lambda)]^2 \right)} \qquad (12.3)$$

El procedimiento para obtener $\hat{\lambda}$ consiste en calcular $L(\lambda)$ para distintos valores de λ. El valor que maximice esta función es el estimador *MV* de la transformación (figura 12.5), y puede obtenerse gráficamente dibujándola por puntos.

Figura 12.5 Estimación gráfica de λ y de un intervalo de confianza



Un contraste de normalidad

Este procedimiento proporciona además intervalos de confianza para el valor de λ y un test de normalidad. En efecto, la distribución del logaritmo del ratio de verosimilitudes es asintóticamente una χ^2, como vimos en el capítulo 10. Por tanto, para el verdadero valor de λ:

$$2[L_{máx}(\lambda) - L(\lambda)] \sim \chi^2(1)$$

y la distribución tendrá un solo grado de libertad al tratarse de un único parámetro.

Fijando un nivel de confianza α, podemos construir un intervalo de confianza para el valor de la función de verosimilitud en el verdadero valor de λ: sea $\chi^2_1(\alpha)$ el valor de la distribución χ^2 con un grado de libertad que deja probabilidad α a la izquierda, entonces:

$$L_{\text{máx}}(\lambda) - L(\lambda) \leq \frac{1}{2}\chi^2_1(\alpha)$$

que implica:

$$L(\lambda) \geq L_{\text{máx}}(\lambda) - \frac{1}{2}\chi^2_1(\alpha)$$

cortando la función $L(\lambda)$ con la ordenada $L_{\text{máx}}(\lambda) - \frac{1}{2}\chi^2_1(\alpha)$ se obtendrán dos valores para el parámetro que definirán un intervalo de confianza para λ. *Si* el Yalor $\lambda = 1$ está incluido en dicho intervalo, aceptaremos la hipótesis de normalidad de los datos con nivel de significación α, mientras que rechazaremos la normalidad en otro caso (figura 12.5). Este contraste es muy potente para detectar asimetría, pero poco eficaz para el apuntamiento.

Ejemplo 12.7

Para ilustrar la utilización de la familia de Box y Cox vamos a estudiar cómo era la distribución de la renta per cápita de las provincias españolas en 1974. Seleccionaremos primero λ gráficamente y después por procedimientos analíticos.

La distribución de la renta provincial por persona está dada en la tabla 12.2.

Observamos en la tabla que hay cinco provincias claramente heterogéneas respecto al resto. La distribución global es asimétrica y sesgada hacia valores bajos, constituyendo las cinco provincias más ricas en grupo aparte. (Estas provincias eran entonces Vizcaya, Madrid, Barcelona, Álava y Guipúzcoa.) La heterogeneidad conduce a que ninguna transformación puede conseguir una apariencia simétrica de todo el conjunto. Por lo tanto, vamos a prescindir de estas cinco provincias excepcionales y analizar el resto como un grupo homogéneo.

La distribución de la tabla 12.2 es sesgada con una cola mayor hacia valores altos, lo que sugiere tomar $\lambda < 1$. Tomando $\lambda = 0$, que corresponde

Tabla 12.2 Renta provincial por persona en miles de pesetas

Intervalo	*f*	Marca de clase	*fr*
49,5 - 59,5	2	55	0,04
59,5 - 69,5	8	65	0,16
69,5 - 79,5	12	75	0,24
79,5 - 89,5	8	85	0,16
89,5 - 99,5	6	95	0,12
99,5 - 109,5	5	105	0,10
109,5 - 119,5	3	115	0,06
119,5 - 129,5	1	125	0,02
129,5 - 139,5	5	135	0,10

al logaritmo, se obtiene la tabla 12.3, que conduce a un histograma aproximadamente simétrico.

Tabla 12.3 Distribución de la renta en logaritmos

Log y	+
3,9 - 4,09	2
4,09 - 4,24	8
4,24 - 4,38	12
4,38 - 4,49	8
4,49 - 4,60	6
4,60 - 4,70	5
4,70 - 4,78	3
4,78 - 4,86	1

Como ejemplo, aplicaremos un contraste de normalidad para ver hasta qué punto los logaritmos de las rentas per cápita de estas 45 provincias españolas eran modelizables mediante una curva normal. Aplicaremos el

contraste χ^2 de Pearson uniendo intervalos para que tengan frecuencias esperadas similares.

Los parámetros de la normal asociada calculados —sin corrección por agrupamiento— de los datos agrupados son:

$$\bar{x} = 4{,}39$$
$$s = 0{,}19$$

Con lo que obtenemos de las tablas de la normal las siguientes frecuencias esperadas:

Intervalo	Fr. ob.	Inter. tipificado	Fr. teór.
3,9 - 4,09	2	$-\infty$ a $-1{,}560$	0,42
4,09 - 4,24	8	$-1{,}560$ a $-0{,}790$	9,24
4,24 - 4,38	12	$-0{,}790$ a $-0{,}076$	11,58
4,38 - 4,49	8	$-0{,}076$ a 0,487	9,55
4,49 - 4,60	6	0,487 a 1,051	7,59
4,60 - 4,70	5	1,051 a 1,560	3,94
4,70 - 4,78	3	1,560 a 1,973	1,58
4,78 - 4,86	1	1,973 a $-\infty$	1,10

Los intervalos tipificados se han calculado mediante la transformación

$$z = \frac{x - \mu}{\sigma}$$

abriendo los dos intervalos extremos para que la suma de las probabilidades de todos ellos sea la unidad.

El contraste de Pearson requiere que las frecuencias esperadas sean al menos tres. Uniendo los dos primeros y los tres últimos, obtendríamos:

Intervalo	Fr. ob.	Fr. teór.
3,9 - 4,24	10	9,66
4,24 - 4,38	12	11,58
4,38 - 4,49	8	9,55
4,49 - 4,60	6	7,59
4,60 - 4,86	9	6,62

$$\chi^2 = \Sigma \frac{(\text{fob} - \text{ft})^2}{\text{ft}} = 1{,}47$$

que corresponde a una probabilidad crítica en la tabla de la χ^2 con dos grados de libertad de 0,45. Por lo tanto, no hay evidencia en los datos para rechazar la hipótesis de normalidad y el ajuste puede considerarse razonablemente bueno.

Para seleccionar λ por máxima verosimilitud, tendremos que:

$$L(\lambda) = -\frac{45}{2} \ln \hat{\sigma}^2(\lambda) + (\lambda - 1) \Sigma f(x_i) \ln x_i$$

Suponiendo que los datos coinciden con las marcas de clase en la tabla 12.2:

$$\Sigma f(x_i) \ln x_i = 2 \ln 55 + 8 \ln 65 + \ldots + \ln 125 = 198{,}42$$

La tabla presenta el cálculo de $L(\lambda)$ con (12.2)

λ	−1,35	−1	−0,9	−0,5	−0,3	−0,1	−0,05
$L(\lambda)$	−131,637	−127,472	−127,102	−126,535	−126,379	−126,319	−126,318

λ	0,01	0,05	0,1	0,6	1,1	1,35	1,6	1,85
$L(\lambda)$	−126,328	−126,336	−126,351	−126,812	−127,802	−128,490	−129,305	−130,237

El máximo de la función es para $\lambda = -0{,}05$ y corresponde a $L(\lambda) = -126{,}318$. Un intervalo de confianza del 95% se obtiene con:

$$L(\lambda) \geq -126{,}318 - \frac{1}{2} 3{,}84 = -128{,}240$$

el intervalo (– 126,32; – 128,24) para $L(\lambda)$ corresponde, según la tabla, al intervalo aproximado de valores de λ:

$$-1{,}2 < \lambda < 1{,}2$$

Este intervalo incluye el valor 1, por lo que se puede aceptar la hipótesis de normalidad en los datos originales al 95%. Sin embargo el máximo de la función es próximo a $\lambda = 0$, lo que nos dice que la mejor transformación de los datos para conseguir simetría es la transformación logarítmica, resultado que habíamos obtenido antes.

Ejercicios 12.1

12.1.1. Contrastar la normalidad de los datos de los experimentos de Michelson y Newcomb (ejercicio 2.15, capítulo 2).

12.1.2. La tabla presenta el número de empates (x) en cada jornada de quinielas durante las temporadas 81-82/82-83. Proponer un modelo probabilístico y contrastar su ajuste.

N.º de clases	0	1	2	3	4	5	6	7
Frecuencia	5	17	25	16	10	4	2	1

12.1.3. Se estudió el tiempo de vida (en horas) de 10 baterías de 9 voltios seleccionadas al azar de la producción, con los resultados siguientes:

28,9; 15,2; 28,7; 72,5; 48,6; 52,4; 37,6; 49,5; 62,1; 54,5

Proponer un modelo de distribución de probabilidad y estudiar su ajuste.

12.1.4. En el ejercicio 12.1.3, se espera que la vida de las baterías siga una distribución exponencial con media 45. ¿Es aceptable la hipótesis?

12.1.5. Al tirar 120 veces un dado se han obtenido los resultados siguientes:

N.º de puntos	1	2	3	4	5	6
Frecuencia	20	14	23	12	26	25

Contrastar la hipótesis de que el dado está equilibrado.

12.1.6. En una encuesta a 100 personas se les preguntó por el número de llaves que llevan habitualmente, obteniendo la siguiente tabla de valores:

N.º de llaves	1	2	3	4	5	6	7	8	9	10
Frecuencia	5	8	16	18	21	12	6	6	4	4

Proponer un modelo de distribución de probabilidad y contrastar su ajuste.

12.1.7. Se han medido 12 valores de una variable física que se supone normal, resultando: 30,2; 30,8; 29,3; 29,0; 30,9; 30,8; 29,7; 28,9; 30,5; 31,2; 31,3 y 28,5. Contrastar que la muestra proviene de una población normal.

12.1.8. En un estudio sobre el tabaco en Andalucía se recogió una muestra de 93 datos y se midió el tiempo de combustión de la hoja de tabaco, obteniéndose la siguiente tabla de frecuencia:

Tiempo de combustión en seg.	0	1	2	3	4	5	6	7	8	9	10	11	12	1 3
Frecuencia	46	15	15	6	3	3	1	1	1	0	0	0	1	1

Encontrar una transformación λ que produzca, aproximadamente, normalidad en esta distribución, contrastando posteriormente ésta.

12.1.9. Los siguientes datos, sacados de una revista del automóvil, proporcionan el coste medio por kilómetro de una muestra de automóviles españoles. Los 108 modelos considerados se han dividido en 11 clases a intervalos de 5 céntimos de euro, quedando:

Frecuencia	5	11	17	26	15	9	8	7	5	3	2	$\Sigma = 108$
Intervalo	20	25	30	35	40	45	50	55	60	65	70	
	25	30	35	40	45	50	55	60	65	70	75	

Proponer un modelo para estos datos y estudiar su ajuste.

12.1.10. Se han tomado datos de estaturas de 502 reclutas españoles de 18 y 19 años en una junta de distrito de Madrid en 1984 con los resultados siguientes:

Frecuencia	6	17	51	119	149	96	48	12	4
Intervalo	150	155	160	165	170	175	180	185	190
en cm.	155	160	165	170	175	180	185	190	195

a) ¿Puede aceptarse que los datos provienen de una distribución normal?
b) Hacer un intervalo de confianza para la media de la población.
c) Sabiendo que para el reemplazo de 1987 la media de los reclutas en toda España fue de 172,9. ¿Hay alguna evidencia de que esta junta de distrito sea atípica?

12.1.11. Se expresan en el siguiente cuadro, agrupados en intervalos, la duración de las películas de cine proyectadas en Madrid entre el 17/II/86 y el 23/II/86.

Frecuencia	4	10	29	31	18	12	7	2	2	2	118 = Total
Intervalo	69,5	79,5	89,5	99,5	109,5	119,5	129,5	139,5	149,5	159,5	
en min.	79,5	89,5	99,5	109,5	119,5	129,5	139,5	149,5	159,5	160,5	

Proponer un modelo para estos datos y contrastar su ajuste.

12.1.12. En la siguiente tabla, se expresa la altura de una muestra de 192 volcanes del mundo agrupados en intervalos, entre 1.000 y 7.000 m.

Frecuencia	19	20	24	26	29	22	16	12	10	8	4	2
Intervalo	1.000	1.500	2.000	2.500	3.000	3.500	4.000	4.500	5.000	5.506	6.000	6.500
en metros	1.500	2.000	2.500	3.000	3.500	4.000	4.500	5.000	5.500	6.000	6.500	7.000

Proponer un modelo de distribución de probabilidad y estudiar su ajuste a los datos.

12.1.13. La tabla proporciona el número de declaraciones de guerra con acciones armadas de más de 50.000 tropas cada año en el período de 1500 a 1931. Proponer un modelo y contrastar su ajuste de datos.

N.º de declaraciones de guerra	0	1	2	3	4	5 o más
Número de años	223	142	48	15	4	0

12.1.14. La tabla muestra el número de soldados muertos en el ejército prusiano por coces de caballos, en diferentes unidades militares de caballería, y es debida a Bortkiewicz. Proponer un modelo y contrastar su ajuste.

N.º de muertos	0	1	2	3	4
Unidades con dichas muertes	109	65	22	3	1

12.2.7 Estimación no paramétrica de densidades

Cuando el objetivo del estudio sea construir un modelo de distribución de probabilidad y los datos rechacen la hipótesis de un modelo concreto podemos estimar directamente la función de densidad a partir de los datos si disponemos de una muestra grande.

El estimador obvio es el histograma: con n datos e intervalos de amplitud $2h$, la estimación del histograma de la función de densidad en el punto x se obtiene haciendo que el área, $2h\,\hat{f}(x)$, sea igual a la frecuencia relativa observada, lo que conduce al estimador:

$$\hat{f}(x) = \frac{1}{2h}\,\frac{(\text{n.º de datos en } x \pm h)}{n} \tag{12.4}$$

Esta estimación tiene la ventaja de la simplicidad y dos inconvenientes principales: (1) es constante dentro del intervalo; (2) es muy dependiente del origen y amplitud de los intervalos, ya que considera únicamente los datos dentro de cada uno, ignorando los adyacentes, por próximos que estén.

Este segundo inconveniente podría resolverse dando cierto peso a los datos en intervalos contiguos al que estimamos, lo que conducirá, además, a una estimación más suave. Sea $n(x \pm h)$ el número de datos en el intervalo $(x \pm h)$ y $n(x + 2h \pm h)$ y $n(x - 2h \pm h)$ los existentes en los intervalos contiguos. Un estimador más suave se obtiene dando cierto peso a la frecuencia relativa de los intervalos contiguos con:

$$\hat{f}(x) = \frac{1}{2hn}\left[\alpha_0 n(x \pm h) + \alpha_1 n(x + 2h \pm h) + \alpha_2 n\,(x + 2h \pm h)\right]$$

donde $\alpha_i > 0$ y $\sum_0^2 \alpha_i = 1$. Tomando por simetría $\alpha_1 = \alpha_2$, y llamando $n(0)$, $n(2h)$, $n(-2h)$ a las frecuencias absolutas de los tres intervalos:

$$\hat{f}(x) = \frac{1}{hn}\left[\frac{\alpha_0}{2}\,n(0) + \frac{\alpha_1}{2}\,(n[2h] + n[-2h])\right] \tag{12.5}$$

donde ahora $\alpha_0 + 2\alpha_i = 1$. El estimador (12.4) corresponde a (12.5) con $\alpha_0 = 1$, $\alpha_1 = 0$. Esta idea puede generalizarse incluyendo el resto de los intervalos con peso decreciente para obtener:

$$\hat{f}(x) = \frac{1}{hn}\left[\frac{\alpha_0}{2}\,n(0) + \frac{\Sigma\alpha_i}{2}\,(n[2ih] + n[-2ih])\right] \tag{12.6}$$

donde los coeficientes α_i verifican $\alpha_0 + 2\Sigma\alpha_i = 1$.

El estimador (12.6) puede ahora aplicarse sin ninguna relación con el histograma: dividimos el rango de valores de la variable en k puntos x_1, ...,

x_k, donde k puede ser tan grande como se quiera, elegimos un valor de h y aplicamos la ecuación (12.6) a cada punto. Por ejemplo, la estimación de $\hat{f}(x_i)$ equivale a construir un histograma con centros de clase:

$$x_i - m2h;\ x_i - (m-1)2h;\ ...;\ x_i - 2h;\ x;\ x_i + 2h;\ ...;\ x_i + m2h$$

y estimar la densidad en el punto x_i aplicando la ponderación simétrica (12.6). Para calcular $\hat{f}(x_{i+1})$ tomamos x_{i+1} como nuevo punto central y aplicamos de nuevo (12.6). Este proceso equivale a calcular la frecuencia absoluta en cada punto dando ciertos coeficientes de ponderación a cada uno de los datos que dependen de su distancia a dicho punto. Escribiendo (12.6) con esta lógica se obtiene:

$$\hat{f}(x) = \frac{1}{hn}\sum_{i=1}^{n} w\left(\frac{x - x_i}{h}\right)$$

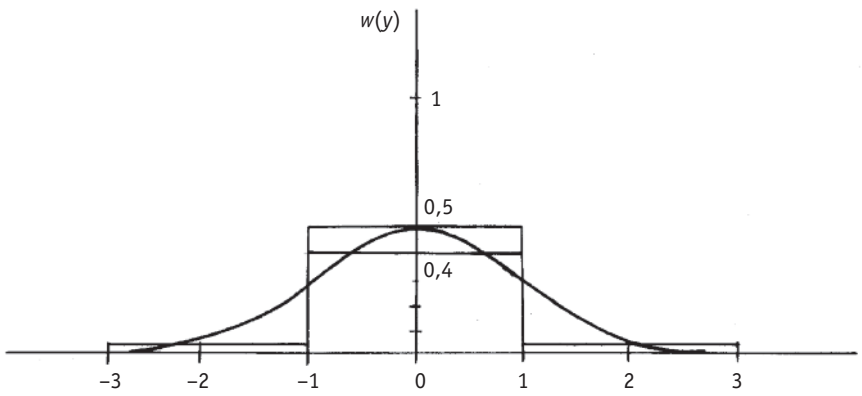
donde w es una función de ponderación que asigna un valor positivo entre cero y uno a cada dato de manera que la suma de todos los pesos sea uno. Por ejemplo, el estimador (12.4) es

$$w(y) = \begin{cases} \frac{1}{2} & |y| \le 1 \\ 0 & \text{en otro caso} \end{cases}$$

mientras que el (12.5), para $0 \le \alpha \le 1$, corresponde a:

$$w(y) = \begin{cases} \alpha/2 & |y| \le 1 \\ (1-\alpha)/2 & 11 < |y| \le 3 \\ 0 & \text{en otro caso} \end{cases}$$

Figura 12.6 Formas posibles de la función de ponderación en estimación de densidades



Ambos estimadores se presentan en la figura 12.6. Es claro que podremos obtener una estimación más suave y precisa de $f(x)$ si utilizamos como $w(y)$ una función continua de ponderación como la indicada en la figura. Por ejemplo, tomando como w la función de densidad normal:

$$\hat{f}(x) = \frac{1}{hn} \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-(x-x_i)^2/2h^2} \tag{12.7}$$

El parámetro clave en esta expresión es h, que representa la semiamplitud de los intervalos que construimos en cada estimación. Como h es también la varianza de la función de ponderación, vamos a dar, en promedio a las observaciones en el intervalo $x \pm h$ un peso entre $(1/\sqrt{2\pi})$ (que equivale a $\phi[0]$, donde ϕ es la función de densidad normal) y $0{,}6/\sqrt{2\pi}$ ($\phi[1]$); un peso entre 0,6 y 0,1 ($\phi[3]$) a los situados en los intervalos adyacentes y prácticamente cero al resto. Este parámetro se determina *anchura de la ventana* o *parámetro de suavizado.* Se denomina función *núcleo* a la función utilizada para determinar las ponderaciones. El estimador (12.7) utiliza como núcleo la función de densidad normal, pero podemos sustituirla por cualquier función que verifique:

$$w(y) \geq 0 \qquad \int_{-\infty}^{\infty} w(y)dy = 1$$

lo que supone que w debe ser una función de densidad. El resultado final depende poco de la elección del núcleo, pero mucho del valor de h.

Elección del parámetro de suavizado

Existen varios métodos para elegir h. Si la distribución que se estima es aproximadamente normal, el parámetro h que minimiza el error cuadrático medio de estimación es (Silverman, 1986)

$$h = 1{,}06 \frac{\sigma}{\sqrt[5]{n}} \tag{12.8}$$

cuando σ es desconocida deberá estimarse con $\hat{s}$ o con un estimador robusto. Esta ventana tiene el inconveniente de ser inadecuada para poblaciones asimétricas. Una solución de compromiso es tomar:

$$h = 0{,}9 \min(\hat{s}_y; RI/1{,}34) \frac{1}{\sqrt[5]{n}} \tag{12.9}$$

donde *RI* es el rango intercuartílico. Este valor de *h* conduce a resultados razonables en una amplia gama de casos. Como regla general, conviene probar con varios valores y comparar los resultados. Cuando *h* es muy grande es fácil comprobar con (12.7) que la distribución estimada será siempre aproximadamente normal.

Ejemplo 12.8

Vamos a estimar una función de densidad a partir de los datos de precios en los países de la OCDE del ejemplo 2.5.

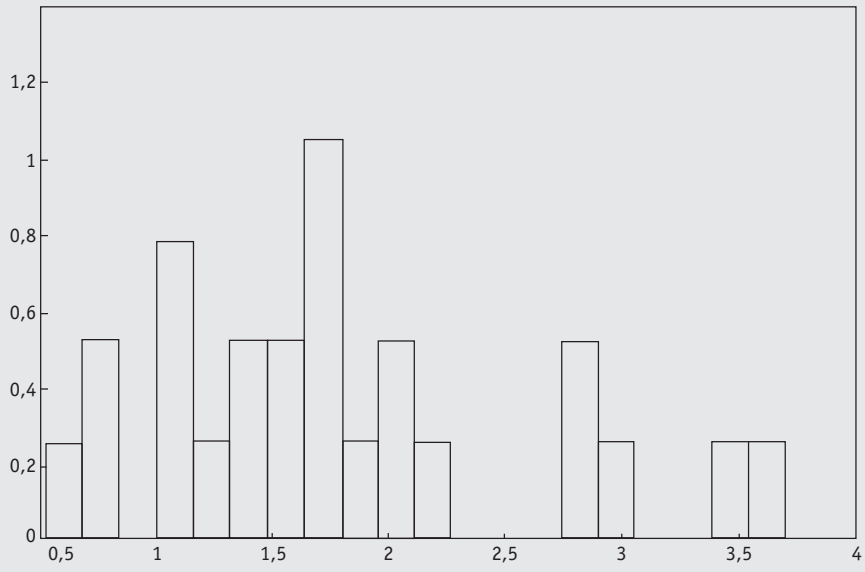
La figura 12.7 presenta el histograma del logaritmo de las observaciones. Admitiendo que forman una muestra de una cierta población cuya densidad tratamos de estimar, como $s_y = 0{,}8118$ (datos en logaritmos) y $RI = 2{,}11 - 1{,}19 = 0{,}92$, la ventana óptima según (12.9) es:

$$h_1 = 0{,}9 \left(\frac{0{,}92}{1{,}34} \right) \frac{1}{\sqrt[5]{24}} = 0{,}32$$

Si los datos fuesen normales, el valor de *h* obtenido con (12.8) es:

$$h_2 = 1{,}06 \cdot \frac{0{,}8118}{\sqrt[5]{24}} = 0{,}45$$

Figura 12.7 Histograma de los datos de precios de la OCDE en logaritmos

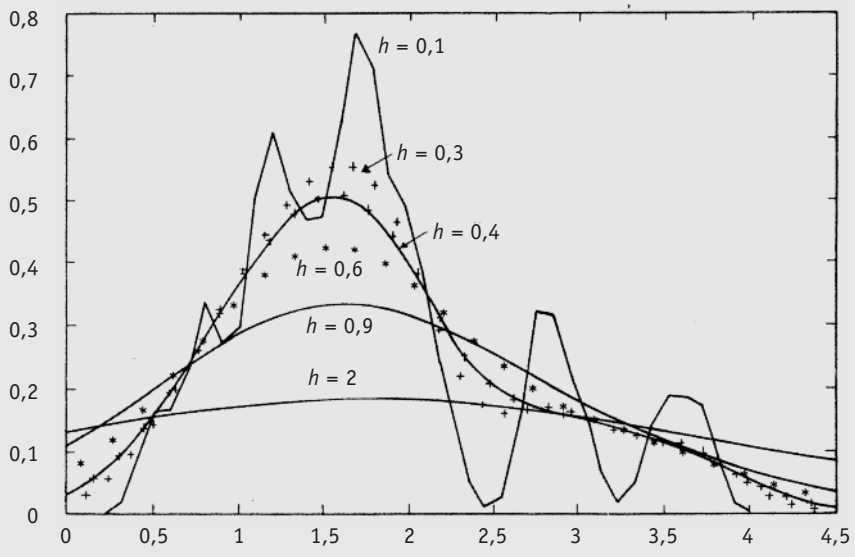


La figura 12.8 muestra la estimación $\hat{f}(x)$ para distintos valores de h. Cuando h es muy pequeño, por ejemplo menor que 0,1, la estimación en cada punto es:

$$\hat{f}(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{2,4} \sum_{i=1}^{n} e^{-1/2(x-x_i)^2/0,1}$$

y como la densidad normal decrece muy rápidamente, la estimación es parecida a la de un histograma con amplitudes pequeñas, hay muy poco suavizado y todos los picos sobresalen muy claramente. Si h es muy grande, por ejemplo 2, la estimación da prácticamente el mismo peso a todos los datos y se obtiene un sobresuavizado (figura 12.8). Se observa que la forma de la distribución aparece bastante claramente para h entre 0,3 y 0,4.

Figura 12.8 Estimación no paramétrica de la función de densidad. Los tamaños de ventana utilizadas sin $h = 0,1$ (línea continua), $h = 0,3$ (+), $h = 0,4$ (+), $h = 0,6$ (*), $h = 0,9$ (continua), $h = 2$ (continua)



12.3 La hipótesis de independencia

12.3.1 Dependencia y sus consecuencias

Cuando las observaciones de la muestra se recogen a lo largo del tiempo o del espacio, es frecuente la aparición de dependencia: las observaciones contiguas tenderán a ser parecidas. Por ejemplo, si observamos las ventas de un producto en *n* días sucesivos, y en general cualquier serie temporal, es frecuente encontrar esta dependencia, que no es esperable cuando los datos se obtienen en el mismo instante temporal por un procedimiento aleatorio.

Cuando las observaciones son dependientes, todas las expresiones utilizadas para las varianzas de los estimadores son erróneas y, por tanto, los intervalos de confianza y los contrastes de hipótesis deducidos a partir de ellos tendrán una confianza o una potencia distinta a la supuesta. El problema es tanto mayor cuanto mayor sea la dependencia. Como ilustración, supongamos que las observaciones muestrales se obtienen en secuencia temporal, y existe una correlación ρ entre cada observación y la precedente, es decir:

$$\text{Cov}(x_i, x_{i+1}) = \rho\sigma^2$$

el coeficiente ρ se denomina *coeficiente de autocorrelación.* Para simplificar, supondremos que no existe relación entre observaciones separadas por dos o más períodos y vamos a estudiar cómo se alteran en este caso las propiedades del estimador más simple: la media muestral. Tendremos:

$$E[\bar{x}] = \frac{1}{n} E[x_1 + ... + x_n] = \frac{n\mu}{n} = \mu$$

y la media muestral sigue siendo un estimador centrado de la media de la población. Su varianza puede escribirse:

$$\text{Var}(\bar{x}) = E\left[\frac{\Sigma x_i}{n} - \mu\right]^2 = \frac{1}{n^2} E[(x_1 - \mu) + ... + (x_n - \mu)]^2$$

desarrollando el cuadrado y tomando esperanzas:

$$\text{Var}(\bar{x}) = \frac{1}{n^2} \Sigma E[(x_i - \mu)^2] + \frac{2}{n^2} \Sigma E[(x_i - \mu)(x_j - \mu)]$$

y como hemos supuesto que todas las covarianzas son nulas con la excepción de las de observaciones contiguas:

$$\text{Var}(\bar{x}) = \frac{\sigma^2}{n} + \frac{2}{n^2}(n-1)\,\rho\sigma^2 = \frac{\sigma^2}{n}\left(1 + \frac{2[n-1]}{n}\,\rho\right)$$

Esta expresión muestra que la variabilidad de la media muestral cuando existe dependencia está afectada por un coeficiente corrector que puede incrementarla, si este coeficiente es mayor que la unidad, o reducirla, si el coeficiente es menor que la unidad. Para ver los valores posibles de este coeficiente impondremos la restricción de que, como la varianza debe ser positiva, el coeficiente debe ser positivo. Esto implica:

$$1 + \frac{2(n-1)}{n}\,\rho \simeq 1 + 2\rho > 0$$

y por lo tanto el coeficiente de autocorrelación debe verificar:

$$\rho > -0{,}5$$

Para analizar el efecto de la dependencia observemos que si la correlación es negativa, el coeficiente corrector es menor que la unidad y la variabilidad de la media muestral es menor que en el caso de datos independientes. Por ejemplo si $\rho = -0{,}45$, el coeficiente multiplicador (suponiendo $[n-1]/n = 1$) es $1 + 2(-0{,}45) = 0{,}1$, que supone que la varianza es sólo un 10% de la existente para observaciones independientes. Por otro lado, si la correlación es positiva, habrá un aumento de la variabilidad. Por ejemplo, si $\rho = 0{,}45$, el coeficiente es $1 + 2(0{,}45) = 1{,}9$, y la varianza se incrementa un 90% respecto al caso de observaciones independientes.

Este cambio en la varianza de $\bar{x}$ modifica las inferencias que hagamos respecto a μ: por ejemplo, si ρ es negativo, el intervalo de confianza para μ partiendo de la hipótesis de independencia será innecesariamente grande, mientras que si ρ es positivo, este intervalo será demasiado pequeño; análogamente, los contrastes de hipótesis respecto a μ, que suponen independencia, serán inválidos. Por ejemplo, para contrastar que la media de una población normal es μ_0, suponiendo observaciones independientes, se utiliza el estadístico:

$$\frac{\bar{x} - \mu_0}{\hat{s}/\sqrt{n}}$$

si existe dependencia positiva, este contraste tenderá a rechazar incorrectamente la hipótesis de que la media es μ_0. En primer lugar $\bar{x} - \mu_0$ no debería compararse con $\hat{s}/\sqrt{n}$, sino con su desviación típica que será considerablemente mayor. En segundo lugar, $\hat{s}^2$ será un mal estimador de σ^2, tendiendo a subestimar la varianza de la población —este resultado es fácil de demos-

trar (ejercicio 12.2.1), pero es intuitivamente claro si tenemos en cuenta que dependencia positiva implica que las observaciones tienden a parecerse entre sí—, con lo que el error será doble.

En resumen, cuando exista dependencia los métodos estudiados no son válidos y podemos obtener fácilmente conclusiones erróneas.

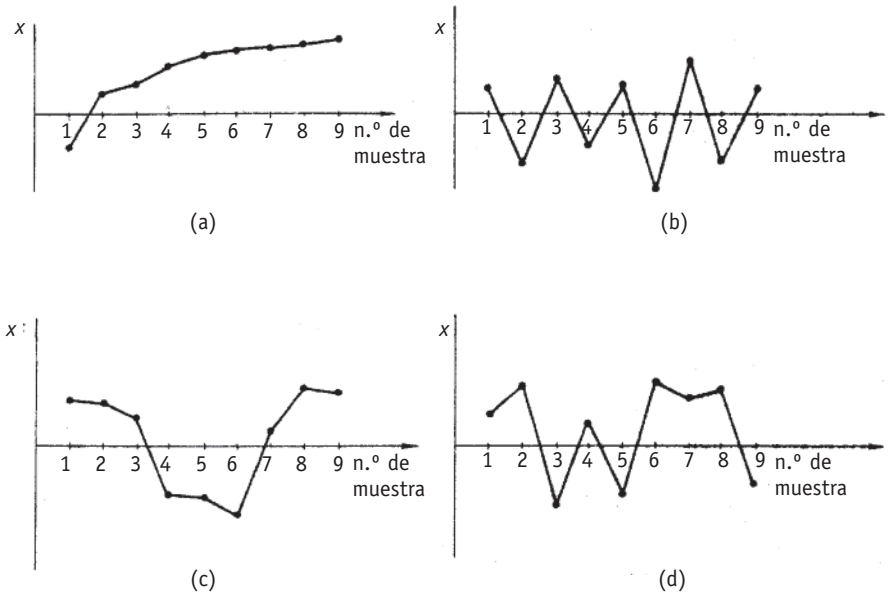
12.3.2 Identificación

Cuando los datos se han obtenido en orden temporal o espacial, conviene siempre dibujarlos en secuencia, para identificar posibles dependencias.

La figura 12.9(a) presenta datos con dependencia positiva, debida a una tendencia creciente; en (b) la dependencia es negativa y se manifiesta en que las observaciones aparecen alternativamente por arriba y debajo de la media; en (c) la dependencia es debida a un comportamiento periódico; finalmente en (d) no se observa una pauta claramente definida.

Vamos a estudiar cómo contrastar esta dependencia.

Figura 12.9 Gráficos de observaciones dependientes



12.3.3 Contraste de rachas

Llamaremos racha a una sucesión de valores por encima o debajo de la mediana. La longitud de la racha es el número de observaciones consecutivas con esta propiedad. Por ejemplo, si los datos son:

31, 52, 80, 45, 62, 50, 37, 72, 75

la mediana es 52 (hay cuatro valores menores y cuatro mayores) y la secuencia, obtenida representando por + los números superiores a la mediana, por – los inferiores y eliminando los iguales a la mediana será:

$$- + - + - - + +$$

existen en total 6 rachas, de longitudes 1, 1, 1, 1, 2, 2. Sea k el número de signos + presentes en la secuencia —que será por hipótesis igual al de signo menos— e igual a $\frac{n-1}{2}$ si n es impar y no hay observaciones repetidas. Puede demostrarse que el número total de rachas en una muestra de n observaciones independientes sigue una distribución aproximadamente normal (si $n > 40$), con parámetros:

$$\mu = k + 1$$

$$\sigma^2 = \frac{k(k-1)}{2k-1}$$

Por tanto, podemos construir un contraste de independencia contando el número de rachas observadas y viendo si este número puede provenir, con una probabilidad razonable, de dicha distribución normal.

Para tamaños muestrales menores que cuarenta, la aproximación anterior no es buena, y conviene utilizar las probabilidades exactas. Los valores críticos de esta distribución se han tabulado (véase tabla 12). Rechazaremos la hipótesis de independencia cuando el número de rachas sea significativamente pequeño, o grande. Por ejemplo, un número de rachas excesivamente grande indica una alta dependencia negativa; un número significativamente pequeño, dependencia positiva.

Ejemplo 12.9

Contrastar si la secuencia:

20, 12, 15, 18, 22, 16, 25, 21, 15, 23, 12, 14

puede considerarse una muestra aleatoria simple de cierta población.

Para obtener la mediana, ordenamos los datos en magnitud:

12, 12, 14, 15, 15, 16, 18, 20, 21, 22, 23, 25,

y la mediana será el valor 17 obtenido mediante:

$$\frac{16 + 18}{2} = 17$$

Sustituyamos ahora cada valor de la secuencia original por un signo + (cuando es mayor que 17) o – (cuando es menor que la mediana); tendremos:

$$+ - - + + - + + - + - -$$

que contiene ocho rachas. No utilizaremos la aproximación normal en este caso ya que el número total de datos es pequeño. En la tabla 12, con $k = 6$, se obtienen los valores críticos:

con $\alpha = 0{,}05$ $\quad 3 \leq r \leq 10$
con $\alpha = 0{,}01$ $\quad 2 \leq r \leq 11$

Por lo tanto, como el valor $r = 8$ obtenido está contenido en el intervalo de aceptación, no hay evidencia suficiente en los datos para rechazar la hipótesis de independencia.

12.3.4 Contraste de autocorrelación

Supongamos una muestra $(x_1, ..., x_n)$ en orden temporal de obtención. Se define el coeficiente de autocorrelación de primer orden, $r(1)$, de la secuencia por:

$$r(1) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(x_{i-1} - \bar{x})}{\Sigma(x_i - \bar{x})^2}$$

Este coeficiente es simplemente el coeficiente de correlación lineal entre las variables $\mathbf{X} = (x_2, ..., x_n)$ e $\mathbf{Y} = (x_1, ..., x_{n-1})$ y es una medida de la relación lineal entre cada observación y la siguiente.

Definiremos el coeficiente de autocorrelación lineal de orden k, $r(k)$, por:

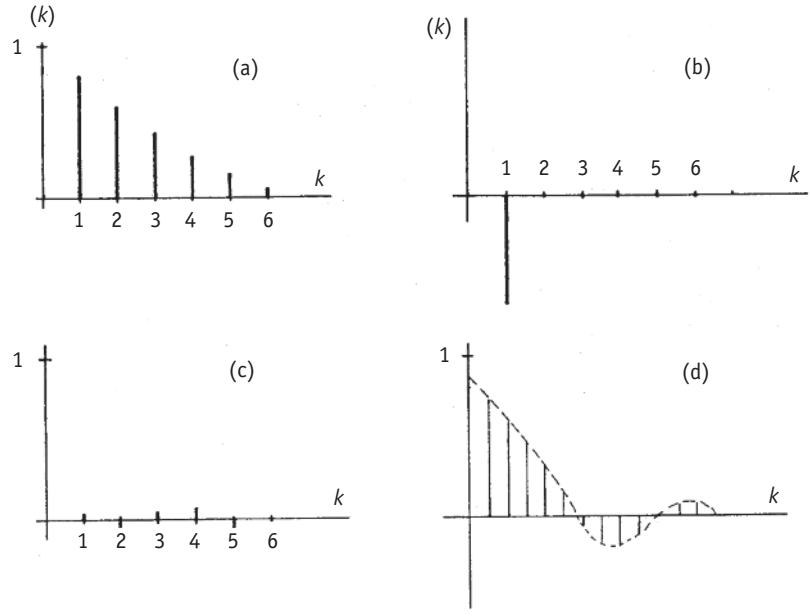
$$r(k) = \frac{\sum_{i=k+1}^{n} (x_i - \bar{x})(x_{i-k} - \bar{x})}{\Sigma(x_i - \bar{x})^2}$$

y representa la relación lineal entre observaciones separadas k posiciones. Observemos que al calcular el coeficiente de orden k no podemos utilizar los primeros k datos en el numerador de la fracción.

Se denomina correlograma o función de autocorrelación a la representación de estos coeficientes de autocorrelación en función del retardo k.

La figura 12.10(a) indica una relación positiva entre cada valor de la secuencia y los anteriores; en (b) la dependencia es negativa, y aparece únicamente en el primer retardo; en (c) todos los coeficientes de autocorrelación son muy pequeños y no hay evidencia, por tanto, de dependencia; en (d) la dependencia es periódica, indicando algún tipo de comportamiento sinusoidal.

Figura 12.10 La función de autocorrelación



En la práctica, si tenemos una muestra no muy grande (entre 20 y 40 observaciones), calcularemos solamente los primeros coeficientes.

Cuando las observaciones son independientes y la población base es normal, los coeficientes de autocorrelación muestrales se distribuyen, aproximadamente, en forma normal, con media cero y varianza $1/n$. Por tanto, podemos considerar significativamente distintos de cero aquellos coeficientes que sean mayores que $2/\sqrt{n}$.

Si disponemos de una muestra grande ($n > 50$) podemos efectuar un contraste conjunto de los primeros coeficientes de autocorrelación. En la hipótesis de independencia:

$$r(k) \to N(0, 1/\sqrt{n})$$

y, por tanto,

$$Q = n \sum_{k=1}^{m} r^2(k)$$

será, aproximadamente, una χ^2 con $m - 1$ grados de libertad. Este test propuesto por Box y Pierce (1970) ha sido mejorado por Ljung y Box (1978) que han demostrado que una aproximación más exacta es considerar el estadístico:

$$Q = n(n + 2) \sum_{k=1}^{m} \frac{r^2(k)}{n - k}$$

que se distribuye como el anterior en la hipótesis de independencia, como una χ^2 con $m - 1$ grados de libertad.

Ejemplo 12.10

El número de matrimonios, redondeado (a miles de personas), en España por meses durante el año 1982 fue según el anuario del INE:

E	F	M	A	M	J	J	A	S	O	N	D
10	9	14	14	17	16	19	24	22	17	13	16

¿Puede considerarse esta secuencia de 20 números como aleatoria?

La simple inspección de la tabla muestra claramente que los datos no son aleatorios y que se producen más matrimonios en verano. Como ilustración, calcularemos los dos primeros coeficientes de autocorrelación muestral. La media y varianza de estos datos son:

$$\bar{x} = 15{,}92 \qquad \hat{s} = 4{,}4$$

$$\sum_{1}^{12} (x_i - \bar{x})^2 = 212{,}9$$

Entonces:

$$\sum_{i=1}^{12}(x_i - \bar{x})(x_{i-1} - \bar{x}) = (10 - 15{,}92)\ (9 - 15{,}92) +$$

$$+ (9 - 15{,}92)(14 - 15{,}92) + \ldots + (13 - 15{,}92)(16 - 15{,}92) = 132{,}90$$

$$r_1 = \frac{132{,}9}{212{,}9} = 0{,}62$$

$$r_2 = \frac{\sum_{i=3}^{12}(x_i - \bar{x})(x_{i-2} - \bar{x})}{\sum_{i=3}^{12}(x_i - \bar{x})^2} =$$

$$= \frac{(14 - 15{,}92)(10 - 15{,}92) + \ldots + (16 - 15{,}92)(17 - 15{,}92)}{212{,}9} =$$

$$= \frac{35{,}75}{212{,}9} = 0{,}17$$

y vemos que parece existir correlación entre observaciones contiguas, ya que el intervalo del 95% para r_1 es, supuesta la independencia:

$$-\frac{2}{\sqrt{n}} < r < \frac{2}{\sqrt{n}}$$

es decir, con $n = 12$

$$-0{,}58 < r < 0{,}58$$

y el valor r_1 obtenido está fuera del intervalo, lo que sugiere que efectivamente existe relación lineal.

Como ilustración calcularemos también el test de Ljung y Box para ambos coeficientes. Entonces:

$$Q = 12 \cdot 14\left(\frac{0{,}62^2}{11} + \frac{0{,}17^2}{10}\right) = 6{,}36$$

que se encuentra más allá del percentil 0,975 de una χ^2 con 1 grado de libertad, lo que conducirá a rechazar la hipótesis de independencia.

12.3.5 Tratamiento de la dependencia

Cuando los datos sean dependientes y constituyan una serie temporal deben analizarse con los métodos que se explican en el último capítulo del segundo tomo.

Ejercicios 12.2

12.2.1. Utilizar los resultados del capítulo 7 para demostrar que la esperanza de la varianza muestral corregida es $\sigma^2(1 - 2/n)$ cuando suponemos que sólo existe dependencia entre observaciones contiguas.

12.2.2. Utilizar el resultado del ejercicio anterior para demostrar que con el esquema de dependencia de la sección 12.3.1 el valor máximo del coeficiente de autocorrelación es $^1/_2$.

12.2.3. Aplicar el contraste de rachas a la serie de matrimonios en España del ejemplo 12.10.

12.2.4. Calcular los coeficientes de autocorrelación para la serie de ventas diarias siguiente:

22, 25, 24, 26, 30, 29, 34, 32, 38, 45, 36, 32, 37, 35, 41.

12.4 La homogeneidad de la muestra

12.4.1 Heterogeneidad y sus consecuencias

Diremos que una muestra es heterogénea cuando todas sus observaciones no han sido generadas por el mismo modelo de distribución de probabilidad. Por ejemplo, algunos datos provienen de un $N(\mu, \sigma)$ y otros de $N(\mu, k\sigma)$ con $k \neq 1$. Las causas más importantes de heterogeneidad son:

1) La población que muestreamos es heterogénea respecto a la variable estudiada. Por ejemplo, tenemos dos clases de elementos y la distribución de la variable es distinta en cada una de ellas. Entonces la heterogeneidad en la muestra representa la existente en la población.
2) La población es homogénea respecto a la variable estudiada, pero en el proceso de muestreo se cometen errores o cambios en las condi-

ciones de medida, consecuencia de los cuales ciertos datos —normalmente una pequeña fracción de la muestra— son heterogéneos (atípicos con el resto).

Vamos a estudiar en primer lugar la heterogeneidad en la población. La influencia de los valores atípicos se aborda en la sección 12.4.5.

12.4.2 Poblaciones heterogéneas: la paradoja de Simpson

Cualquier población real puede ser heterogénea: entre otros factores, las personas difieren por sexo, educación, procedencia social; las empresas por sector de actividad, localización, tipo de fabricación, los elementos fabricados en un proceso por turno de trabajo, tipo de proceso utilizado y clase de materia prima.

Al estudiar estadísticamente la distribución de una variable en una población cuyos elementos pueden clasificarse en grupos, es importante tener en cuenta que la distribución puede ser distinta en los distintos grupos. Si las diferencias son pequeñas, podemos ignorarlas e incluirlas en el error experimental. Por ejemplo, si mezclamos poblaciones normales con la misma varianza, y con medias que diferen menos de una desviación típica, obtenemos de nuevo una población aproximadamente normal. Si las proporciones de cada subpoblación en el total se distribuyen aproximadamente de forma normal, de nuevo obtenemos poblaciones normales. Esto justifica que las distribuciones de muchas medidas físicas sean normales (estatura, peso, etc.) aunque difieran un poco en distintos estratos de la población. Sin embargo, otras variables físicas (la longitud del pie, por ejemplo) son claramente distintas para hombres y mujeres y uniendo ambos sexos obtenemos una distribución bimodal. En este caso tratar la población humana como homogénea puede llevar a errores considerables, como veremos a continuación.

La tabla 12.4a presenta la proporción de admitidos en una universidad clasificados por sexo. Si suponemos homogeneidad y que los 4.000 estudiantes implicados son una muestra aleatoria de la población de estudiantes pasados y futuros, concluiríamos que hay una diferencia significativa en la admisión a favor de las mujeres.

La tabla 12.4b presenta estos datos desagregados por facultades. Se observa que las tres facultades muestran discriminación a favor de los hombres. Por tanto, las conclusiones de los datos divididos en subpoblaciones más homogéneas son opuestas a los datos agregados. Este fenómeno se conoce como la *paradoja de Simpson.*

La explicación de esta paradoja es la siguiente: supongamos una población de N elementos que puede dividirse en k subpoblaciones distintas con N_1, ..., N_k elementos. Sean n_1, ..., n_k el número de elementos de cada subpo-

Tabla 12.4a Admisiones a una universidad por sexo

	Solicitudes	Admisiones	Proporción
Mujeres	2.000	1.136	56,80%
Hombres	2.000	955	47,75%

Tabla 12.4b Admisiones por facultades y sexo

		Solicitudes	Admisiones	Proporción
LETT	*Mujeres*	800	560	70%
	Hombres	300	225	75%
ING	*Mujeres*	200	36	18%
	Hombres	700	140	20%
ECO	*Mujeres*	1.000	540	54%
	Hombres	1.000	590	59%

blación con la característica que se desea estudiar. Entonces, la proporción total de elementos con dicha característica es:

$$p_T = \frac{n_1 + ... + n_k}{N_1 + ... + N_k} = \frac{\Sigma n_i}{N_T}$$

y llamando $p_i = n_i/N_i$ a la probabilidad en cada subpoblación y $f_i = N_i/N_T$ a su frecuencia relativa en el total, podemos escribir:

$$\boxed{p_T = \Sigma f_i p_i} \qquad (12.10)$$

que indica que la probabilidad total es una media ponderada de las probabilidades parciales. Si comparamos dos sucesos A y B, es posible que p_A sea

mayor que p_B en todas las poblaciones pero que en el total ocurra el suceso contrario, que es la paradoja de Simpson. La ecuación (12.10) es también válida para una muestra siendo p_T la proporción total observada y p_i las muestrales en cada subpoblación cuya representación relativa en la muestra es f_i.

Este fenómeno puede ocurrir igualmente con variable continuas. Llamando $\bar{x}_i$ a las medias muestrales de la subpoblación i ($i = 1, ..., k$) y f_i a su representación relativa en la muestra, es inmediato comprobar que:

$$\boxed{\bar{x} = \Sigma f_i \bar{x}_i} \qquad (12.11)$$

fórmula que generaliza la (12.10). La tabla 12.5 presenta un ejemplo de esta paradoja para variables continuas. Supongamos que se desea comparar la vida media de dos lotes de bombillas. Se toma una muestra de diez bombillas del proveedor p_1 y resulta una vida media de 520 horas, que sube a 700 horas para el proveedor p_2. Un estudio más detallado conduce a comprobar que hay tres tipos de bombillas con características distintas y que la duración de cada tipo es mayor en el caso del provedor p_1, contrariamente a lo que parecía a primera vista. El error proviene de la distancia composición de las dos muestras.

Tabla 12.5 Comparación de dos muestras heterogéneas

	Media muestral	Subpoblaciones		
		A	B	C
p_1	$\bar{x}$ = 520 h (n = 10)	400 (6)	600 (3)	1.000 (1)
p_2	$\bar{y}$ = 700 h (n = 10)	300 (2)	500 (2)	900 (6)

12.4.3 Identificación de la heterogeneidad: contraste de Wilcoxon

La característica común a una población heterogénea es una alta variabilidad y un bajo coeficiente de curtosis. En consecuencia, siempre que la muestra presente este rasgo conviene comprobar si podemos dividirla en muestras independientes homogéneas y contrastar si existen diferencias entre ellas. La heterogeneidad para datos normales puede provenir de las medias, las varianzas o ambos parámetros y, en caso de duda, conviene aplicar los tests correspondientes estudiados en el capítulo 10.

En esta sección vamos a presentar un contraste general para comprobar si dos muestras independientes provienen de una misma población continua: el contraste de Wilcoxon. El segundo tomo del libro está dedicado a estudiar cómo comprobar y medir el efecto de otras variables sobre una variable de interés que es continua, por lo que todas las técnicas que allí se exponen pueden utilizarse para investigar la homogeneidad de una muestra de variables normales.

Cuando la variable respuesta es un atributo, el análisis de la homogeneidad puede realizarse con técnicas que expondremos en la sección siguiente.

Contraste de Wilcoxon

Supongamos dos muestras $(x_1, ..., x_n)$, $(y_1, ..., y_m)$ independientes de una variable continua (o una muestra que subdividimos en dos submuestras). Se trata de contrastar que ambas muestras provienen de la misma población. Unimos las dos muestras para formar una muestra única, y ordenamos las observaciones de menor a mayor. Por ejemplo:

$$x_6 < y_3 < y_7 < x_1 \ldots < y_{14} < y_{25}$$

Llamaremos *rango* de un dato al orden que ocupa en esta ordenación. Por ejemplo x_6 tendrá rango 1, y_3 rango 2, y_{25} rango $n + m$. Sea R_x la suma de los rangos de las x y R_y la de las y. Es indiferente considerar uno u otro ya que su suma, que es la suma de los $n + m$ rangos, es constante. Llamando $N = n + m$:

$$\sum_{1}^{n+m} i = \left(\frac{1 + n + m}{2} \right) (m + n) = N \frac{N + 1}{2}$$

Si las distribuciones de x e y son idénticas, cualquiera de las ordenaciones posibles tiene la misma probabilidad. Por tanto, el rango de una observación cualquiera, $r(x_i)$, tomará los valores 1, ..., N con la misma probabilidad, $1/N$.

Entonces:

$$E[r(x_i)] = \sum_{i=1}^{N} i \frac{1}{N} = \left(\frac{N + 1}{2} \right)$$

y como este resultado es válido para cualquier observación:

$$E[R_x] = E[r(x_1) + ... + r(x_n)] = nE[r(x_i)]$$

$$\boxed{E[R_x] = n \frac{(N + 1)}{2}}$$

Puede comprobarse que, en la hipótesis de igualdad entre ambas poblaciones:

$$\text{Var}[R_x] = \frac{n \cdot m}{12}(N+1)$$

y la variable R_x es aproximadamente normal si el tamaño muestral no es muy pequeño ($n, m > 5$).

El contraste consiste en calcular $E[R_x]$ y $DT[R_x]$ y construir un intervalo de confianza para el número esperado muestral de rachas. Rechazaremos la hipótesis cuando este número sea muy pequeño o muy grande.

Este contraste se denomina de Wilcoxon y también de Mann-Withney, y es casi tan potente como el contraste *t* de Student para las medias en la hipótesis de normalidad (95,5% de eficiencia relativa asintótica) y mucho más potente para otras distribuciones. Un resultado importante debido a Lehmann es que la eficiencia relativa asintótica de este contraste respecto al de la *t* de Student no puede ser menor del 86,4%.

Empates

En el análisis anterior se ha supuesto que no hay dos observaciones iguales. Aunque esto ocurrirá en teoría, ya que suponemos modelos continuos, en la práctica es posible encontrar valores idénticos en ambas muestras. La solución más simple para tratar con estos empates es realizar dos contrastes: en el primero siempre que hay dos observaciones idénticas asignamos a la X el rango mayor; en el segundo hacemos lo contrario. Si ambos contrastes conducen a la misma conclusión, el problema está resuelto; en caso contrario los datos no son concluyentes.

Por ejemplo, si $X = (8, 10, 15, 15, 20)$, $Y = (8, 12, 15, 20, 22)$, los dos contrastes serán:

a) El orden de los datos asignando a X el rango mayor es (un * indica dato de la muestra X):

   8, 8*, 10*, 12, 15, 15*, 15*, 20, 20*, 22

   y el rango de X es $2 + 3 + 6 + 7 + 9 = 27$

b) El orden asignando a Y el rango mayor es:

   8*, 8, 10*, 12, 15*, 15*, 15, 20*, 20, 22

   y el rango de X es ahora $23(1 + 3 + 5 + 6 + 8)$

En consecuencia, los dos contrastes se harán con $R_x = 27$ y con $R_x = 23$. Como:

$$E[R_x] = 5 \cdot \frac{11}{2} = 27{,}5$$

$$DT[R_x] = \sqrt{\frac{25}{12} \cdot 11} = 4{,}79$$

es claro que en ambos casos se aceptará la hipótesis de homogeneidad: en el primero se obtiene casi el valor esperado y en el segundo un valor a menos de una desviación típica de la media.

Ejemplo 12.11

Se ha tomado una muestra aleatoria de tamaño 15 de las calificaciones de dos profesores que corrigen un mismo examen. Los resultados obtenidos son:

A: 7,6; 5,8; 3,4; 7,4; 7,5; 4,5; 8,7; 7,0
B: 2,9; 7,1; 7,7; 6,5; 5,3; 3,8; 4,8

¿Califican los profesores de forma distinta?

Ordenemos las dos muestras. Un * indica la muestra *A*:

2,9; 3,4*; 3,8; 4,5*; 4,8; 5,3; 5,8*; 6,5; 7,0*; 7,1; 7,4*; 7,5*; 7,6*; 7,7; 8,7*

El rango de la muestra *A* es:

$$R_A = 2 + 4 + 7 + 9 + 11 + 12 + 13 + 15 = 73$$

y el de la *B*:

$$R_B = 1 + 3 + 5 + 6 + 8 + 10 + 14 = 47$$

y su suma es:

$$15 \cdot 8 = 73 + 47 = R_A + R_B$$

En la hipótesis de igualdad la variable R_A será aproximadamente normal, con parámetros:

$$E[R_A] = 8 \cdot \frac{16}{2} = 64$$

$$DT[R_A] = \sqrt{\frac{7 \cdot 8}{12}}\sqrt{16} = 8{,}64$$

Por tanto el valor 73 corresponde a:

$$Z = \frac{73 - 64}{8{,}64} = 1{,}04$$

y no hay evidencia suficiente de que las calificaciones son distintas.

12.4.4 Análisis de tablas de contingencia

El análisis de tablas de contingencia es un procedimiento general para investigar la homogeneidad de poblaciones cualitativas. En síntesis, el método consiste en comparar las frecuencias observadas para cada atributo dentro de cada clase con las esperadas por un modelo que suponga homogeneidad en todas las clases o categorías.

Vamos a comprobar que los contrastes para atributos estudiados se reducen a casos particulares de este método general: comparar las frecuencias observadas con las esperadas según la ecuación del contraste χ^2 de la sección 12.2.2. Esta relación permitirá generalizar el contraste de comparación de dos muestras, para comparaciones entre k muestras.

Contrastes binomiales y contraste χ^2

Comenzaremos con el contraste más simple para un atributo. Este contraste es H_0: $p = p_0$, frente a H_1: $p \neq p_0$. Para muestras grandes, el contraste es:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} \tag{12.12}$$

donde $\hat{p}$ es la proporción observada en una muestra de tamaño n. Este contraste resulta también del análisis siguiente: dispongamos en una tabla las frecuencias observadas y las previstas por la hipótesis:

	Esperadas	Observadas
A	np_0	$n\hat{p}$
$\bar{A}$	nq_0	$n\hat{q}$

si aplicamos el contraste x^2:

$$X^2 = \Sigma \frac{(\text{Observadas} - \text{Esperadas})^2}{\text{Esperadas}}$$

como tenemos únicamente dos frecuencias, el estadístico resultante será, si la hipótesis es cierta, una χ^2 con un grado de libertad, y su expresión es:

$$X^2 = \frac{(np_0 - n\hat{p})^2}{np_0} + \frac{(nq_0 - n\hat{q})^2}{nq_0} = \frac{n(\hat{p} - p_0)^2}{p_0 q_0}$$

que es el cuadrado de (12.12), y ambos contrastes son idénticos.

Análogamente, el contraste de igualdad de las proporciones en dos muestras utiliza el estadístico:

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_0\hat{q}_0}{n_1} + \frac{\hat{p}_0\hat{q}_0}{n_2}}} \tag{12.13}$$

donde:

$$\hat{p}_o = \frac{n_1}{n_1 + n_2}\hat{p}_1 + \frac{n_2}{n_1 + n_2}\hat{p}_2 \tag{12.14}$$

La tabla 12.6 presenta las frecuencias observadas y las esperadas (entre paréntesis) de acuerdo con la hipótesis de igualdad de las dos muestras.

Tabla 12.6 Frecuencias observadas y esperadas (entre paréntesis) en un contraste de dos muestras

	Muestra 1	Muestra 2	Conjunta
Frecuencia de A	$n_1\hat{p}_1$ $(n_1\hat{p}_0)$	$n_2\hat{p}_2$ $(n_2\hat{p}_0)$	$n_1\hat{p}_1 + n_1\hat{p}_2$ $[(n_1 + n_2)\hat{p}_0]$
Frecuencia de $\bar{A}$	$n_1\hat{q}_1$ $(n_1\hat{q}_0)$	$n_2\hat{q}_2$ $(n_2\hat{q}_0)$	$n\hat{q}_1 + n_2\hat{q}_2$ $[(n_1 + n_2)\hat{q}_0]$
TOTALES	n_1	n_2	$n_1 + n_2$

con lo que el estadístico (x^2) resulta ser, en este caso:

$$X^2 = \frac{n_1(\hat{p}_1 - \hat{p}_0)^2}{\hat{p}_0} + \frac{n_2(\hat{p}_2 - \hat{p}_0)^2}{\hat{p}_0} + \frac{n_1(\hat{q}_1 - \hat{q}_0)^2}{\hat{q}_0} + \frac{n_2(\hat{q}_2 - \hat{q}_0)^2}{\hat{q}_0}$$

y utilizando que $(\hat{p}_i - \hat{p}_0)^2 = (\hat{q}_i - \hat{q}_0)^2 = n_i^2(\hat{p}_1 - \hat{p}_2)^2/(n_1 + n_2)^2$ se obtiene que:

$$X^2 = \frac{(\hat{p}_1 - \hat{p}_2)^2}{\hat{p}_0\hat{q}_0\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

que es de nuevo el cuadrado del contraste de proporciones (12.13). El estadístico resultante tiene un grado de libertad, ya que tenemos dos frecuencias independientes y hemos estimado el parámetro (12.14) para calcular las frecuencias teóricas. Este número coincide con el número de casillas independientes cuando fijamos las sumas en los márgenes de la tabla.

Contrastes de homogeneidad

El método anterior puede extenderse sin dificultad para analizar cualquier muestra de atributos que puede clasificarse en categorías. Supongamos que estudiamos el número de veces que aparecen k posibles atributos mutuamente excluyentes (A_1, ..., A_k) en una muestra de n elementos, y que los elementos pueden clasificarse en c grupos distintos (G_1, ..., G_c) dando lugar a una tabla de contingencia con $k \times c$ casillas (tabla 12.7).

Tabla 12.7 Una tabla de contingencia general

	G_1	.....	G_j	.....	G_c	
A_1	f_{11}	.....	f_{1j}	.....	f_{1c}	$f_{1.}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
A_i	f_{i1}	.....	f_{ij}	.....	f_{ic}	$f_{i.}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
A_k	f_{k1}	.....	f_{kj}	.....	f_{kc}	$f_{k.}$
TOTALES	$f_{.1}$	.....	$f_{.j}$	.....	$f_{.c}$	n

Vamos a estudiar el contraste:

H_0: los grupos no influyen y la muestra es homogénea
H_1: hay diferencias entre los grupos.

Si H_0 es cierta, las mejores estimaciones de las probabilidades de cada atributo son:

$$\hat{p_i} = P(A_i) = \frac{f_{i.}}{n}$$

Por tanto, la frecuencia esperada en cada fila será el resultado de multiplicar esta probabilidad estimada de la fila, si no hay diferencias, por el número de elementos en cada grupo:

$$E_{ij} = \text{frecuencia esperada } (ij) = f_{.j}\frac{f_{i.}}{n}$$

En resumen, la frecuencia esperada en cada casilla es el producto de las frecuencias marginales dividido por n.

Si la hipótesis de homogeneidad es cierta, el estadístico:

$$X^2 = \sum_{i=1}^{k}\sum_{j=1}^{c}\frac{(f_{ij} - E_{ij})^2}{E_{ij}} = \sum_{i=1}^{k}\sum_{j=1}^{c}\frac{(nf_{ij} - f_{.j}f_{i.})^2}{nf_{i.}f_{j.}}$$

será una χ^2 con tantos grados de libertad como tengan las frecuencias de la tabla fijadas las marginales. Este número es $(k-1)\times(c-1)$.

Este análisis se generaliza sin dificultad para tablas de cualquier dimensión. Por ejemplo, si clasificamos con tres criterios tendremos una tabla tridimensional, y si sus dimensiones son $k \times c \times r$, el estadístico (X^2) tendrá ahora $(k-1)\times(c-1)\times(r-1)$ grados de libertad. El problema de aumentar la dimensión es que las restricciones estudiadas en la sección (12.2.2) para que este estadístico se distribuya como una X^2 deben mantenerse: la frecuencia esperada de cada casilla debe ser al menos tres. En tablas grandes, sin embargo, puede admitirse que algunas casillas tengan una frecuencia esperada no menor que 0,5.

El contraste de homogeneidad es en definitiva un contraste de independencia entre dos criterios de clasificación de las observaciones y se utiliza con frecuencia con este objetivo.

Ejemplo 12.12

La tabla siguiente proporciona los alumnos matriculados por sexos en una muestra de facultades de ciencias económicas y empresariales en el curso 85/86. ¿Es distinta la proporción de mujeres en las distintas universidades?

	Alcalá	Alicante	Barna A.	Barna C.	Mad. A.	Mad. C.
Hombres	1.394	1.558	2.142	6.854	5.583	10.458
Mujeres	515	594	609	2.140	2.309	3.919
TOTAL	1.909	2.152	2.751	8.994	7.847	14.377

Suponiendo homogeneidad en las facultades, la mejor estimación de la proporción de mujeres es:

$$\hat{p}_M = \frac{515 + \ldots + 3.919}{1.909 + \ldots + 14.377} = \frac{10.086}{38.030} = 0{,}265$$

lo que conduce a la siguiente tabla de frecuencias esperadas.

	AH	A	BA	BC	MA	MC	Totales
Hombres	1.403	1.581	2.021	6.609	5.766	10.564	27.944
Mujeres	506	571	730	2.385	2.081	3.813	10.086
Totales	1.909	2.152	2.751	8.994	7.847	14.377	38.030

donde la primera fila se obtiene multiplicando la fila de totales por 0,735 (27.944/38.030), y la segunda por 0,265. El estadístico será:

$$\chi^2 = \frac{9^2}{1.403} + \frac{23^2}{1.581} + \ldots + \frac{106^2}{3.813} = 0{,}05 + 0{,}334 + \ldots + 24{,}9 + 2{,}94 = 100{,}9$$

y, si la hipótesis es cierta, corresponde a una χ^2 con 5 grados de libertad. Como este valor no puede venir de dicha distribución, rechazamos la hipótesis de homogeneidad y concluimos que hay diferencias significativas entre las universidades. En las universidades de Barcelona la proporción de mujeres es algo menor (0,23) que en el resto (0,27).

Ejercicios 12.3

12.3.1. A una muestra de 200 personas de ambos sexos se les dio a probar margarina y mantequilla y se les pidió indicasen su preferencia, con los resultados de la tabla. ¿Hay diferencia entre los sexos?

	Margarina	**Mantequilla**
Hombres	42	58
Mujeres	65	35

12.3.2. La tabla siguiente presenta las calificaciones medias de un grupo de estudiantes en dos asignaturas. ¿Hay diferencias entre ellas?

	Calificación		
Asignatura	Baja	Media	Alta
A	15	18	7
B	2	22	23

12.3.3. En una encuesta entre estudiantes sobre su creencia en la percepción ultrasensorial se encontraron los datos siguientes. ¿Hay diferencias entre las creencias según el tipo de estudios escogidos?

	Creen totalmente	**A medias**	**En absoluto**	
Ingeniería	30	50	20	100
Económicas	50	109	11	170
Humanidades	48	93	9	150
	128	243	40	420

12.3.4. Los datos siguientes muestran las frecuencias resultantes de dos medidas en la fabricación de 99 piezas. ¿Varían ambas medidas de forma independiente?

	B(mm)		
A (mm)	40-42	43-45	46-49
10-15	40	36	2
16-20	0	14	7

12.4.5 El efecto de datos atípicos

Un caso de heterogeneidad importante se produce cuando una pequeña fracción de la muestra (entre el 1 y el 10%) aparece como atípica, debido a errores de medición o codificación de los datos, a cambios en los instrumentos de medida y, en general, a alteraciones en el proceso de recogida de datos. El efecto de esta heterogeneidad puede ser muy grave: supongamos una muestra de n observaciones de una población (μ_0, σ_0) donde, por error, uno de los datos proviene de una población distinta $(\mu_1, k\sigma_0)$. Entonces:

$$E[\bar{x}] = \frac{n-1}{n}\mu_0 + \frac{1}{n}\mu_1$$

La media muestral será un estimador sesgado. Si μ_1 es mucho mayor que μ_0, y n no es muy grande, la media muestral tendrá un error alto como estimador de μ_0. Su varianza será:

$$\text{Var}(\bar{x}) = \frac{\sigma_0^2}{n}\left(1 + \frac{k^2-1}{n}\right)$$

y si k^2 es grande, la varianza puede ser arbitrariamente grande. Por tanto, una única observación muy atípica puede alterar todas las propiedades de los estimadores. En general, los estimadores estudiados son muy poco robustos ante la heterogeneidad. Este hecho fue ilustrado en la sección 7.5.5 al presentar los estimadores robustos.

Existen dos filosofías básicas para el tratamiento de valores atípicos. La primera es modificar el proceso de estimación para que los parámetros no se vean afectados por estos valores anormales. La segunda es identificarlos mediante un test e indagar las causas que los motivan, eliminándolos de la muestra si se confirma su heterogeneidad.

El proceso de estimación puede a su vez modificarse de dos maneras:

1) Suponiendo un modelo más general que permita la aparición de valores atípicos.
2) Utilizando estimadores robustos, como los presentados en el capítulo 7 (véase también apéndice 2D, donde se presentan los M - estimadores).

En el primer caso suele suponerse que los datos se generan con alta probabilidad $(1 - \alpha)$ (α pequeña) por un modelo $f(\mu, \sigma)$, pero a veces, con probabilidad α, provienen de otra distribución $f(\mu, k\sigma)$ con $k > 1$. La distribución que genera las observaciones es, por tanto, la mezcla:

$$f(x, \mu, \sigma, k, \alpha) = (1 - \alpha)f(x; \mu, \sigma) + \alpha f(x; \mu, k\sigma)$$

El modelo depende ahora de cuatro parámetros que pueden estimarse por el método de máxima verosimilitud. Las ecuaciones resultantes son no lineales y deben resolverse iterativamente. Los estimadores obtenidos, $\hat{\mu}$ y $\hat{\sigma}$, no estarán contaminados por los valores extremos (véase el apéndice 12D).

Una muestra con alto coeficiente de apuntamiento sugiere, en general, una distribución mezclada del tipo anterior. En efecto, la varianza de la variable será:

$$\text{Var}(x) = (1 - \alpha)\sigma^2 + \alpha k^2\sigma^2 = \sigma^2(1 + \alpha[k^2 - 1])$$

y es fácil comprobar que:

$$E[(x - \mu)^4] = 3\sigma^4(1 + \alpha[k^4 - 1])$$

Por tanto, el coeficiente de apuntamiento de la variable normal contaminada será:

$$CAp = 3\,\frac{[1 + \alpha(k^4 - 1)]}{[1 + \alpha(k^2 - 1)]^2}$$

Por ejemplo, si $\alpha = 0{,}01$ y $k = 4$, $Cap = 8{,}05$ y el apuntamiento de la distribución mezclada o mixta será mucho mayor que el de la distribución normal.

Es interesante señalar que cuando la heterogeneidad en la muestra proviene de la existencia de dos poblaciones distintas que se mezclan, el coeficiente de apuntamiento en general disminuye en lugar de aumentar. Por ejemplo, supongamos que los datos son una mezcla al 50% de dos poblaciones normales con distinta media y la misma desviación típica. Supongamos que, sin pérdida de generalidad, la primera es $N(-a, 1)$ y la segunda $N(a, 1)$, de manera que la mezcla al 50% tiene media cero. Es fácil comprobar que entonces su varianza es $1 + a^2$, y que el coeficiente de apuntamiento viene dado por

$$CAp = \frac{3 + 6a^2 + a^4}{[1 + a^2]^2}$$

el apuntamiento es siempre menor que 3, y si a es grande tiende al mínimo valor posible 1. En consecuencia, el coeficiente de apuntamiento es muy útil para detectar heterogeneidad: si es grande, sugiere la presencia de valores atípicos, y si es pequeño, sugiere la presencia de distribuciones mezcladas o, lo que es equivalente, de un grupo grande de atípicos homogéneos que son distintos del resto de los datos muestrales. En la sección siguiente se presenta un contraste de valores atípicos.

12.4.6 Test de valores atípicos

En la hipótesis de que los datos son normales, el test más simple para verificar si el valor máximo (o mínimo) de una muestra de tamaño n puede considerarse heterogéneo es:

$$q_m = \text{máx}\left[\frac{x_{(n)} - \bar{x}}{\hat{s}}, \frac{\bar{x} - x_{(1)}}{\hat{s}}\right] = \text{máx}\left|\frac{x_i - \bar{x}}{\hat{s}}\right|$$

donde $x_{(n)}$ es el valor máximo y $x_{(1)}$ el mínimo de la muestra; q_n es, por tanto, la máxima distancia entre una observación y la media. La distribución de q_n en la hipótesis de que toda la muestra proviene de una distribución normal se ha tabulado. Algunos valores importantes se dan en la tabla 12.8.

Tabla 12.8 Valores críticos para el test de valores anómalos

n	5	6	7	8	9	10	12	15	20
$\alpha = 5\%$	1,71	1,89	2,02	2,13	2,21	2,29	2,41	2,55	2,71
$\alpha = 1\%$	1,76	1,97	2,14	2,28	2,38	2,48	2,63	2,81	3,00

Este test debe utilizarse solamente para muestras pequeñas donde se sospeche la presencia de una sola observación atípica. Para muestras medianas de poblaciones normales se obtiene un test más conveniente, que tiene en cuenta la presencia simultánea de varios datos atípicos, calculando el coeficiente de apuntamiento por:

$$CAp = \frac{\Sigma(x_i - \bar{x})^4}{n\hat{s}^4}$$

cuya distribución en el muestreo para muestras homogéneas y normales se ha tabulado (tabla 12.9). Admitiremos la presencia de varios valores atípicos cuando el apuntamiento de la distribución sea significativamente mayor que el de la normal.

Tabla 12.9 Valores críticos para el test de apuntamiento

n	5	10	15	20	25	50	75	100	200	500
$\alpha = 5\%$	2,9	3,9	4,1	4,1	4,0	3,99	3,87	3,77	3,57	3,37
$\alpha = 1\%$	3,1	4,8	5,1	5,2	5,0	4,88	4,59	4,39	3,98	3,60

Para n grande, los valores de la tabla 12.9 están calculados teniendo en cuenta que entonces el *CAp* de una variable normal, como hemos visto en la sección 12.2.4, se distribuye como $N(3, \sqrt{24/n})$. Por tanto:

$$\frac{CAp - 3}{\sqrt{24/n}} \sim N(0, 1)$$

Por ejemplo, con $n = 500$

$$\frac{CAp - 3}{\sqrt{24/500}} = \frac{CAp - 3}{0{,}219} \sim N(0, 1)$$

luego el valor crítico para un contraste unilateral al 95% se obtendrá de:

$$\frac{CAp - 3}{0{,}219} < 1{,}645$$

que implica:

$$CAp < 3 + 0{,}219 \cdot 1{,}645 = 3{,}37$$

que es el valor indicado en la tabla 12.9. Los valores menores de 3 para tamaños muestrales pequeños provienen de que entonces el valor de *CAp* está acotado por el tamaño muestral.

12.4.7 Tratamiento de los atípicos

Cuando se encuentran valores atípicos en la muestra, hay dos posibles explicaciones. La primera es que los atípicos corresponden a errores de medición y en consecuencia deben eliminarse de la muestra para calcular los estimadores de los parámetros del modelo. La segunda es considerar que no se trata de errores de medición, sino que el modelo generador de los datos tiene colas pesadas y puede generar con cierta probabilidad valores que se alejan mucho del centro de los datos. En este segundo caso se ilustra en el apéndice 12D que el estimador *MV* de una distribución con colas pesadas implica dar menos peso a las observaciones extremas para estimar el centro de los datos. Este resultado es general: el estimador *MV* de una distribución con colas pesadas debe dar un peso más pequeño a las observaciones alejadas, y ese peso tiende a cero para observaciones muy extremas. En consecuencia, sea cual sea la hipótesis que finalmente escojamos para explicar la aparición de los atípicos, el comportamiento que debemos seguir es eliminar —o darles un peso muy pequeño— a las observaciones extremas para estimar la media de los datos.

La estimación de la variabilidad es sin embargo distinta en ambas hipótesis. Si los atípicos son errores de medida que pueden eliminarse en el futuro, los atípicos deben eliminarse para estimar la variabilidad. Sin embargo, si son valores generados por la distribución, deben tenerse en cuenta para calcular la variabilidad de los datos.

12.5 Resumen del capítulo

En este capítulo se ha abordado el problema fundamental de comprobar las hipótesis básicas de construcción del modelo: la forma de la distribución, la independencia y la homogeneidad. La heterogeneidad es probablemente la más importante, ya que puede afectar a la base misma del proceso de inferencia: los datos no representan la población objetivo. Dado que la heterogeneidad puede manifestarse de muchas formas distintas, conviene siempre estudiar con detalle los datos muestrales utilizando las herramientas descriptivas del capítulo 2 y los contrastes expuestos en este capítulo.

La autocorrelación es de nuevo un problema de los datos que afecta gravemente al proceso de inferencia. Sin embargo, es fácil de identificar y existen técnicas estadísticas adecuadas para modelar datos dependientes.

Finalmente, la forma de la distribución no suele ser un problema grave siempre que los datos sean homogéneos. Los procedimientos estudiados son globalmente válidos, aunque dejan de ser óptimos.

La tabla 12.10 resume los métodos principales estudiados en este capítulo.

12.6 Lecturas recomendadas

Los contrastes de ajuste se tratan en casi todos los libros de estadística básica de la bibliografía y, con especial detalle, en los dedicados a métodos no paramétricos. Buenas referencias son Conover (1999) y Breiman (1973). La estimación de densidades, en Silverman (1986) y Nadaraya (1989). El análisis de tablas de contingencia, en Fienberg (2007) y Everitt (1992); el estudio de valores atípicos, en Barnett y Lewis (1994), y los métodos robustos, en Hampel *et al.* (1986) y Haber (1981). Los contrastes no paramétricos en Conover (1999), Mosteller y Rourker (1973) y Lehmann (2006), entre otros.

Tabla 12.10 Estadísticos principales introducidos en el capítulo

1. Contrastes de ajuste	Nombre	Estadístico	Distribución
General	*ji*-cuadrado	$\Sigma \frac{(\text{Observadas-Esperadas})^2}{\text{Esperadas}}$	χ^2_{n-p-1}
V. continuas	Kolmogorov-Smirnov	$\text{Sup}\ \lvert F_n(x) - F(x)\rvert$	Tabla 8
Normalidad	Shapiro y Wilk	$\frac{1}{ns^2}\sum_{1}^{n} a_{j,n}[x_{(n-j-1)} - x_{(j)}]$	Tablas 10 y 11
	K-S-Lilliefors	$\text{Sup}\ \lvert F_n(x) - F(x)\rvert$	Tabla 9
	Asimetría	$\sqrt{n}\ CA/\sqrt{6}$	$N(0, 1)$
	Apuntamiento	$\sqrt{n}\ (CAp - 3)\ \sqrt{24}$	$N(0, 1)$
	Conjunto	$n[CA^2/6 + (CAp - 3)^2/24]$	χ^2_2
	Transformación de Box-Cox	$x^{(\lambda)} = \frac{(x+m)^\lambda - 1}{\lambda}$	
Estimación de densidades	Métodos núcleo	$f(x) = \frac{1}{hn}\Sigma \frac{1}{\sqrt{2\pi}} e^{-1/2(x-x_i)^2/h^2}$	

Tabla 12.10 Estadísticos principales introducidos en el capítulo *(continuación)*

2.	Independencia	Rachas	x = n.º de rachas	$x \sim N\left[k+1; \dfrac{k(k+1)}{2k-1}\right]$ y tabla 12
		Autocorrelación	$Q = n(n+2)\sum_{1}^{m} \dfrac{r^2(k)}{n-k}$	χ^2_{m-1}

3.	Homogeneidad	Nombre	Estadístico	Distribución
	Dos muestras	Wilcoxon	Rango de los datos de una muestra	$R_x \sim N\left[n\dfrac{(N+1)}{2}; \sqrt{\dfrac{nm}{12}(N+1)}\right]$
	Atributos	Tablas de contingencia ($r \times c$)	$\Sigma \dfrac{(\text{Observadas-Esperadas})^2}{\text{Esperadas}}$	$\chi^2\ (r-1)x(c-1)$
	Un dato atípico población normal	Desviación máxima estudentizada	máx $\left\lvert \dfrac{x_i - \bar{x}}{\hat{s}} \right\rvert$	Tabla 12.8
	Varios datos atípicos poblaciones normales	Coeficiente de apuntamiento	$\sqrt{n}\ (CAp-3)/\sqrt{24}$	Tabla 12.9 y $N(0, 1)$

Apéndice 12A: El contraste χ^- de Pearson

Puede deducirse un contraste de ajuste utilizando el método de la razón de verosimilitudes. Sean O_1, ..., O_k las frecuencias esperadas. Las hipótesis a contrastar son:

$$H_0: E[O_i] = E_i = np_{oi}$$
$$H_1: E[O_i] \neq E_i$$

donde P_{o1}, ..., P_{ok} son las probabilidades de cada clase especificadas por el modelo. Sea p_i la probabilidad verdadera de obtener un valor en la clase *i*. La función de verosimilitud será:

$$\ell(p_1, ..., p_k) = p_1^{O1} \ldots p_k^{Ok}$$

y en logaritmos:

$$L(\mathbf{p}) = \Sigma O_i \ln p_i$$

estando definida en el conjunto de valores definido por $0 \leq p_i \leq 1$; $\Sigma p_i = 1$. Para aplicar el contraste de razón de verosimilitudes debemos obtener los estimadores *MV*. Como la función está restringida por la ecuación $\Sigma p_i = 1$, tendremos que utilizar los multiplicadores de Lagrange. Llamando ϑ al parámetro de Lagrange, la función a maximizar es:

$$M(\mathbf{p}, \vartheta) = \Sigma O_i \, lnp_i - \vartheta(\Sigma p_i - 1)$$

El máximo verificará:

$$\frac{\partial M}{\partial p_i} = \frac{O_i}{p_i} - \vartheta = 0; \qquad \frac{\partial M}{\partial \vartheta} = 0 = \Sigma p_i - 1$$

Por tanto:

$$O_i = \vartheta p_i; \qquad \Sigma O_i = n = \vartheta \Sigma p_i = \vartheta$$

es decir:

$$\hat{p_i} = \frac{O_i}{n}$$

que será el estimador *MV*. En dicho punto, tendremos:

$$L(\mathbf{p}) = \Sigma O_i \ln \frac{O_i}{n}$$

El contraste de verosimilitudes es:

$$TV = 2 \ln \lambda = 2[L(\mathbf{p}) - L(\mathbf{p}_0)] = 2\left(\Sigma O_i \ln \frac{O_i}{n} - \Sigma O_i \ln p_{oi}\right) = 2\Sigma O_i \ln \frac{O_i}{E_i}$$

Esta expresión es asintóticamente equivalente al contraste χ^2 de Pearson. Para verlo, expresemos O_i/E_i como $1 + (O_i - E_i)/E_i$; entonces, desarrollando en serie de Taylor, como $\ln(1 + x) \simeq x - x^2/2 +$ términos menores que x^2:

$$\ln\left(\frac{O_i}{E_i}\right) \simeq \frac{O_i - E_i}{E_i} - \frac{1}{2}\left(\frac{O_i - E_i}{E_i}\right)^2$$

despreciando términos de orden superior. Por tanto, sustituyendo en la expresión de $2\ln\lambda$, escribiendo O_i como $(O_i - E_i + E_i)$ y operando se obtiene que:

$$2 \ln \lambda \simeq \Sigma \left(\frac{O_i - E_i}{E_i}\right)^2$$

que es el contraste χ^2.

Este contraste resulta directamente como un test del gradiente, expuesto en el apéndice 10B. En efecto, los componentes del vector gradiente en la función de verosimilitud M, que tiene en cuenta la restricción $\Sigma p_i = 1$, son:

$$\frac{\partial M}{\partial p_i} = \frac{O_i}{p_i} - \vartheta = \frac{O_i}{p_i} - n = \frac{O_i - np_i}{p_i}$$

y las segundas derivadas y su esperanza para construir la matriz de información esperada:

$$\frac{\partial^2 M}{\partial p_i^2} = -\frac{O_i}{p_i^2}\ ;\ \frac{\partial^2 M}{\partial p_i p_j} = 0\ ;\ E\left[\frac{\partial^2 M}{\partial p_i^2}\right] = -\frac{np_i}{p_i^2} = -\frac{n}{p_i}$$

Por tanto, particularizando para H_0 que establece $np_i = E_i = np_{i0}$:

$$\left(\frac{\partial M}{\partial p_i}\right)_{poi} = n\frac{(O_i - E_i)}{E_i}\ ;\ E\left(\frac{\partial^2 M}{\partial p_i^2}\right)_{poi} = -\frac{n^2}{E_i}$$

y el test resultante se obtiene aplicando la fórmula (5C.2). Como la matriz de segundas derivadas es diagonal, la forma cuadrática resultante se reduce a:

$$TG = n^2\Sigma\left(\frac{O_i - E_i}{E_i}\right)^2 \left(\frac{n^2}{E_i}\right)^{-1} = \Sigma\frac{(O_i - E_i)^2}{E_i}$$

que es el contraste χ^2.

Apéndice 12B: Deducción del contraste de Shapiro y Wilk

Un método gráfico conveniente, para juzgar respecto a la normalidad de una muestra de pequeño tamaño ($n < 30$), es el diagrama probabilístico normal, cuyo fundamento es el siguiente: Supongamos la muestra ordenada:

$$x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$$

En la hipótesis de que estos valores provienen de una distribución normal con media μ y varianza σ^2, los valores estandarizados:

$$\frac{x_{(1)} - \mu}{\sigma} \leq \frac{x_{(2)} - \mu}{\sigma} \leq \ldots \leq \frac{x_{(n)} - \mu}{\sigma}$$

serán una muestra ordenada de una población $N(0, 1)$, cuyos valores esperados están tabulados. Sea $C_{i,n}$ el valor esperado del término que ocupa el lugar i en una muestra de tamaño n de una población normal, es decir:

$$E\left[\frac{x_{(i)} - \mu}{\sigma}\right] = C_{i,n}$$

Entonces:

$$E[x_{(i)}] = \mu + \sigma C_{i,n}$$

Por tanto, el gráfico de $x_{(i)}$ respecto a $C_{i,n}$ será, aproximadamente, una recta cuya ordenada en el origen estimará μ y su pendiente σ.

La expresión general de los coeficientes $C_{i,n}$ es complicada, pero pueden aproximarse por:

$$C_{i,n} = \Phi^{-1}\left(\frac{i - 3/8}{n + 1/4}\right)$$

donde Φ es la función de distribución normal estándar. Esta expresión muestra que los valores esperados de los estadísticos ordenados son simétricos. En efecto, en una $N(0, 1)$, por simetría:

$$\Phi(-z) + \Phi(z) = 1$$

Sustituyendo en la expresión de $C_{i,n}$ es inmediato comprobar que para n par:

$$\Phi(C_{i,n}) + \Phi(C_{n+1-i,n}) = 1 \qquad i = 1, ..., \frac{n}{2}$$

Por tanto, concluiremos que $C_{in} = -C_{n+1-i,n}$ y sólo necesitamos calcular la mitad de los coeficientes $C_{i,n}$, ya que los otros los obtendremos por simetría; además, la suma

$$\sum_{i=1}^{n} C_{i,n}$$

es siempre cero, ya que cuando n es impar el valor esperado del término central es cero.

Vamos a construir un test de normalidad a partir del diagrama probabilístico-normal midiendo el ajuste de los puntos a una recta por el cuadrado del coeficiente de correlación lineal entre ambas variables, dado por:

$$r^2 = \frac{[\Sigma x_{(i)} C_{i,n}]^2}{ns^2(\Sigma C_{i,n}^2)}$$

donde no se ha restado en el numerador el producto de las medias de las variables porque los coeficientes $C_{i,n}$ tienen media cero. El test resultante puede escribirse utilizando la simetría de los coeficientes $C_{i,n}$ en la forma de Shapiro y Wilk:

$$w = \frac{1}{ns^2}\left[\sum_{j=1}^{h} a_{j,n}(x_{[n-j+1]} - x_{[j]})\right]^2 = \frac{A^2}{ns^2}$$

donde s^2 es la varianza muestral, h es $n/2$ si n es par y $(n - 1)/2$ si es impar y los coeficientes $a_{j,n}$ se obtienen con:

$$a_{j,n} = \frac{|C_{i,n}|}{\Sigma C_{i,n}^2}$$

Shapiro y Wilk han tabulado los valores exactos de $a_{j,n}$.

Apéndice 12C: Selección gráfica de la transformación

Representaremos por x_p el percentil de orden p de los datos, es decir, aquel valor tal que al ordenar los n datos de la muestra ocupa el lugar $[np]$, supuesto que dicho número es entero. Con datos simétricos x_p y x_{1-p} deben estar a la misma distancia de la media o mediana. Por tanto, al transformar se desea que:

$$x_p^{(\lambda)} - x_{0,5}^{(\lambda)} = x_{0,5}^{(\lambda)} - x_{1-p}^{(\lambda)}$$

y si la transformación λ consiguiese exactamente simetría:

$$M^{(\lambda)} = x_{0,5}^{(\lambda)} = \frac{x_p^{(\lambda)} + x_{1-p}^{(\lambda)}}{2}$$

y la mediana debería ser igual a la media de los percentiles simétricos, para cualquier p.

En general, si transformamos los datos, cualquier percentil $x_p^{(\lambda)}$ puede desarrollarse en serie de Taylor alrededor del valor $M^{(\lambda)}$ como sigue:

$$x_p^{(\lambda)} \simeq M^{(\lambda)} + \left[\frac{dx_p^{(\lambda)}}{d\lambda}\right]_M (x_p - M) + \frac{1}{2}\left[\frac{d^2x_p^{(\lambda)}}{d\lambda^2}\right]_M (x_p - M)^2$$

y como:

$$\frac{dx_p^{(\lambda)}}{d\lambda} = x_p^{\lambda-1}$$

$$\frac{d^2x_p^{(\lambda)}}{d\lambda^2} = (\lambda - 1)x_p^{\lambda-2}$$

tendremos que:

$$x_p^{(\lambda)} \simeq M^{(\lambda)} + M^{\lambda-1}(x_p - M) + \frac{\lambda - 1}{2} M^{\lambda-2}(x_p - M)^2$$

Sustituyendo esta expresión para x_p y x_{1-p} en la ecuación de $M^{(\lambda)}$ y operando, tenemos que:

$$\frac{x_p + x_{1-p}}{2} - M \simeq (1 - \lambda)\frac{(x_p - M)^2 + (x_{1-p} - M)^2}{4M} \qquad (12C.1)$$

Esta ecuación indica que si los datos transformados son aproximadamente simétricos y calculamos para distintos valores de p los términos:

$$y(p) = \frac{x_p + x_{1-p}}{2} - M$$

$$z(p) = \frac{(x_p - M)^2 + (x_{1-p} - M)^2}{4M}$$

los puntos $y(p)$, $z(p)$ deberán estar en una recta que pasa por el origen y tiene pendiente $1 - \lambda$. Por tanto, λ puede estimarse aproximadamente como sigue:

1) Ordenar la muestra y seleccionar varios valores (entre 4 y 8) de p. Normalmente tomaremos p de manera que recojamos bien el comportamiento de los extremos que es donde habrá más variabilidad y cuidando que np sea entero. Una selección indicativa es tomar para p valores del orden de (0,25), (0,15), (0,10), (0,05), (0,03), (0,01) o números parecidos.
2) Calcular la mediana M y para cada valor de p los percentiles x_p y x_{1-p}. Introducir estos tres números en la ecuación (12C.1) para determinar un valor de λ.
3) El paso anterior proporciona tantos valores de λ como valores de p. Como estimador de λ se tomará la mediana de los valores obtenidos, que es un estimador robusto, poco afectado por algún valor atípico.
4) Alternativamente a (3) pueden llevarse los puntos $y(p)$, $z(p)$ a un gráfico y determinar a simple vista una recta. Su pendiente nos dará el valor $1 - \lambda$.

El procedimiento anterior es simple y proporciona una estimación rápida de un valor aproximado para λ. Sin embargo, este procedimiento es poco preciso, y cuando sea posible es más conveniente utilizar el método de máxima verosimilitud.

Apéndice 12D: Estimadores robustos iterativos

Normal-contaminada

Suponiendo que cada elemento de la muestra $\mathbf{X} = (x_1, ..., x_n)$ viene de una distribución $N(\mu, \sigma)$ con probabilidad $(1 - \alpha)$, y de $N(\mu, k\sigma)$ con probabilidad α, y suponiendo α y k constantes fijas, la verosimilitud es:

$$\ell(\mu \mid \mathbf{X}) = \prod_{i=1}^{n} \left[(1-\alpha) \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2} + \alpha \frac{1}{\sigma k\sqrt{2\pi}} e^{-1/2\left(\frac{x_j-\mu}{k\sigma}\right)^2} \right]$$

tomando logaritmos y derivando el soporte para obtener el estimador *MV*:

$$\frac{dL(\mu \mid \mathbf{X})}{d\mu}=0=\Sigma\frac{(1-\alpha)(x_i-\mu)\,e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}+k^{-1}\alpha\left(\frac{x_j-\mu}{k^2}\right)e^{-1/2\left(\frac{x_j-\mu}{k\sigma}\right)^2}}{(1-\alpha)\,e^{-1/2\left(\frac{x_j-\mu}{\sigma}\right)^2}+\alpha k^{-1}\,e^{-1/2\left(\frac{x_j-\mu}{\sigma k}\right)^2}} \quad (12D.1)$$

Sea $P(M_1 \mid x_i)$ la probabilidad de que la observación x_i venga del modelo $M_1 = N(\mu, \sigma)$. Según el teorema de Bayes esta probabilidad es:

$$P(M_1 \mid x_i)=\frac{P(x_i \mid M)P(M)}{P(x_i)}=\frac{(1-\alpha)\,e^{-1/2\left(\frac{x_j-\mu}{\sigma}\right)^2}}{(1-\alpha)\,e^{-1/2\left(\frac{x_j-\mu}{\sigma}\right)^2}+\alpha k^{-1}\,e^{-1/2\left(\frac{x_j-\mu}{\sigma k}\right)^2}} \quad (12D.2)$$

Con este resultado, la ecuación (12D.1) puede escribirse:

$$\Sigma(x_i-\mu)\,[P(M_1|x_i)+k^{-2}P(M_2|x_i)]=0 \quad (12D.3)$$

y como:

$$P(M_2|x_i)=1-P(M_1|x_i)=p_i$$

llamando p_i a la probabilidad de que la observación sea atípica, (12D.3) se reduce a:

$$\Sigma(x_i-\mu)\,[1-p_i(1-1/k^2)]=0 \quad (12D.4)$$

Según esta ecuación, cuando $p_i \simeq 0$ la observación *i* tiene un peso igual a la unidad, mientras que si $p_i \simeq 1$ tiene peso $1/k^2$. En casos intermedios (12D.4) asigna a cada dato una ponderación entre 0 y k^{-2}. Llamando:

$$w_i=1-p_i(1-1/k^2) \quad (12D.5)$$

a estos coeficientes de ponderación, (12D.1) puede finalmente escribirse como:

$$\Sigma(x_i-\mu)w_i=0 \quad (12D.6)$$

cuya solución es:

$$\hat{\mu}=\frac{\Sigma x_i w_i}{\Sigma w_i} \quad (12D.7)$$

La ecuación (12D.7) pone de manifiesto que la estimación de $\hat{\mu}$ puede realizarse por el siguiente proceso iterativo: fijar α (normalmente entre 0,05 y 0,2) y k (entre 3 y 5) y obtener un estimador inicial robusto de σ. Con estos tres parámetros y suponiendo un valor inicial para μ (por ejemplo la mediana) calculamos las probabilidades (12D.2) y los pesos (12D.5) que sustituidos en (12D.6) conducen a un nuevo estimador de $\hat{\mu}$. Este valor se utiliza de nuevo ahora para recalcular las probabilidades (12D.2) y el proceso se repite hasta obtener convergencia.

Por supuesto el método puede mejorarse incorporando una segunda ecuación de máxima verosimilitud para estimar σ. La estimación simultánea de los cuatro parámetros es difícil —la función de verosimilitud suele tener muchos máximos— y es mejor estimar (μ, σ) condicionadas a (α, k) y luego probar la sensibilidad de la estimación a distintos valores de estos estimadores.

Un procedimiento para estimar mezclas es utilizar el algoritmo EM. Véase por ejemplo Peña (2001) para su aplicación al caso multivariante.

M-estimadores

Se denominan M-estimadores a los resultantes de modificar el método de máxima verosimilitud para que sea robusto a desviaciones de la normalidad. Para un parámetro de centralización μ, este método maximiza

$$L(\mu) = k - {}^{1}/_{2} \Sigma \left(\frac{x_i - \mu}{\sigma} \right)^2$$

que conduce a:

$$L'(\mu) = 0 = \Sigma \left(\frac{x_i - \hat{\mu}}{\sigma} \right) \qquad (12D.8)$$

y proporciona el estimador $\hat{\mu} = \bar{x}$. Este estimador no es robusto porque la función $(x_i - \mu)/\sigma$ no está acotada, y un valor cualquiera puede tener un peso ilimitado en la estimación.

Una solución intuitivamente sensata es generalizar (12D.7) y escribir la ecuación (12D.8) como:

$$\Sigma \left(\frac{x_i - \mu}{\sigma} \right) w_i = 0 \qquad (12D.9)$$

donde los w_i son ciertos coeficientes de ponderación a determinar del tipo:

$$w_i = \begin{cases} 1, & \text{si } \left|\frac{x-\mu}{\sigma}\right| \leq b \\ \text{decreciente hacia cero si } \left|\frac{x-\mu}{\sigma}\right| > b \end{cases}$$

es decir, cuando las observaciones se encuentran en un intervalo «razonable» (±1,7 por ejemplo) no se modifican (reciben peso unidad), pero a medida que se alejan de dicho valor reciben un peso decreciente, que tiende a cero para valores muy extremos. Esto equivale a sustituir la ecuación (12D.8) por:

$$\Sigma \Psi \left(\frac{x_i - \mu}{\sigma} \right) = 0 \qquad (12D.10)$$

donde la función Ψ está relacionada con los pesos w_i (12D.9) por:

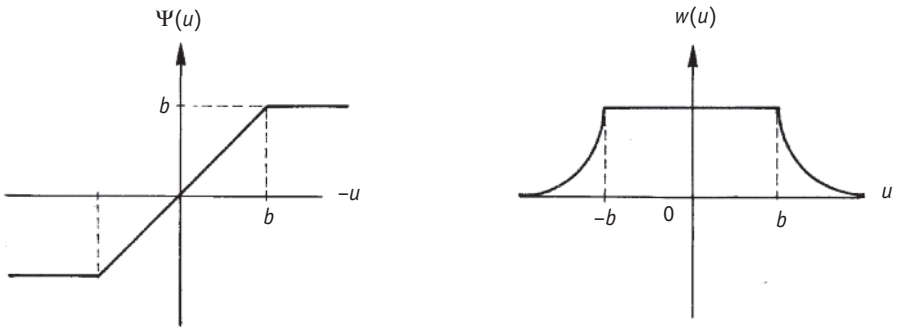
$$\Psi_i \left(\frac{x_i - \mu}{\sigma} \right) = \left(\frac{x_i - \mu}{\sigma} \right) w_i \left(\frac{x_i - \mu}{\sigma} \right) \qquad (12D.11)$$

El M-estimador queda definido al indicar el valor de Ψ o w, ya que, por (12D.11), una función determina la otra.

$$\Psi(u) = u;\ |u| \leq b;\ \Psi(u) = b,\ |u| > b.\ w(u) = \Psi(u)/u$$

La figura 12.11 presenta ambas funciones para un esquema de ponderación muy utilizado, sugerido por Huber (1983). Por supuesto, existen otros muchos esquemas posibles, y según (12D.6) cada uno es óptimo para un tipo especial de contaminación. En la práctica el resultado es poco sensible a cualquier función de ponderación que «descuente» las observaciones atípicas. El procedimiento de cálculo de un M-estimador es el siguiente:

Figura 12.11 Función de ponderación de Huber



1) Comenzar con un estimador robusto inicial de μ, $\hat{\mu}^{(0)}$. (Por ejemplo, la mediana o la media recortada al 20%.)
2) Calcular un estimador robusto de σ, como $\hat{\sigma}$ = MEDA/0,675.
3) Utilizando $\hat{\mu}^{(0)}$ y $\hat{\sigma}$ calcular las ponderaciones w_i. Por ejemplo, con la ecuación de la figura (12.11) fijando el valor de b (por ejemplo 1,7), los pesos se calculan con:

$$w_i^{(0)} = \begin{cases} 1, & \text{si } \left| \dfrac{x_i - \hat{\mu}^{(0)}}{\hat{\sigma}} \right| \leq b \\ \dfrac{b}{\left| \dfrac{x_i - \hat{\mu}^{(0)}}{\hat{\sigma}} \right|}, & \text{si } \left| \dfrac{x_i - \hat{\mu}^{(0)}}{\hat{\sigma}} \right| > b \end{cases} \tag{12D.12}$$

4) Con las ponderaciones $w_i^{(0)}$ calcular un nuevo estimador $\hat{\mu}^{(1)}$ mediante:

$$\hat{\mu}^{(1)} = \frac{\Sigma x_i w_i^{(0)}}{\Sigma w_i^{(0)}} \tag{12D.13}$$

5) Partiendo del nuevo estimador obtenido en (4), repetir el cálculo de los pesos en (3) y la estimación del parámetro en (4) e iterar hasta obtener convergencia, es decir, hasta que:

$$|\hat{\mu}^{(u+1)} - \hat{\mu}^{(i)}| < \varepsilon$$

para un valor pequeño ε prefijado.

Para estudiar las propiedades de los estimadores así obtenidos, supongamos para simplificar y sin pérdida de generalidad que el verdadero valor de μ es cero, y desarrollemos la ecuación (12D.10) en serie en un entorno del verdadero valor ($\mu = 0$). Entonces:

$$\Sigma\Psi(x_i - \hat{\mu}) \simeq \Sigma[\Psi(x_i) - \Psi'(x_i)\hat{\mu}] = 0$$

que implica:

$$\hat{\mu} = \frac{\Sigma\Psi(x_i)/n}{\Sigma\Psi'(x_i)/n} \tag{12D.14}$$

cuando n sea grande, el numerador puede escribirse como:

$$\bar{y} = \frac{\Sigma y_i}{n}$$

donde $y_i = \Psi(x_i)$. Si la distribución de x es simétrica, se verifica que:

$$E[y] = \int_{-\infty}^{\infty} \Psi(x) f(x) dx = 0$$

ya que $\Psi(x) = -\Psi(-x)$. Por tanto, para tamaños muestrales grandes $\bar{y}$ será próximo a cero. El denominador tiende a una constante para tamaños muestrales altos, ya que, por ejemplo con la función de Huber de la figura 12.11:

$$\Psi'(x) = \begin{cases} 1 & |x| \leq b \\ 0 & |x| > b \end{cases}$$

y $E[\Psi'(x)] = P(|x| \leq b) = \text{cte}$. En consecuencia, para tamaños muestrales grandes:

$$\frac{\Sigma \Psi'(x_i)}{n} \simeq E[\Psi'(x)] \qquad (12D.15)$$

Estos resultados indican que el estimador $\hat{\mu}$ será centrado para grandes muestras. Su varianza se obtiene sustituyendo (12D.15) en (12D.14):

$$\text{Var}(\hat{\mu}) = \frac{1}{n} \frac{E[\Psi(x)^2]}{E[\Psi'(x)^2]} \qquad (12D.16)$$

donde las esperanzas están tomadas respecto a la distribución verdadera que ha generado los datos. Por tanto, conocida la función Ψ, pueda obtenerse con (12D.16) la varianza asintótica del M-estimador.

Cuarta parte

Control de calidad

13. Control de calidad

W. Edwards Deming (1900-1993)
Estadístico estadounidense. Ha hecho contribuciones fundamentales a la teoría del muestreo, el control de calidad y a la dirección de empresas. La importancia de su enseñanza en Japón después de la Segunda Guerra Mundial sobre métodos estadísticos para la calidad ha sido reconocida por el premio Deming, que concede anualmente el gobierno japonés a la empresa que haya destacado más en la mejora de la calidad.

13.1 Introducción

En todo proceso productivo o administrativo se puede definir una medida de la calidad de sus resultados. La medida de calidad puede ser cualitativa: una visita comercial puede resultar o no en un contrato, un documento administrativo puede contener o no errores y un elemento fabricado puede ser o no defectuoso. Con frecuencia la medida de calidad es continua: el tiempo requerido para proporcionar un servicio (en servicio al cliente), la longitud en una pieza, el peso de un producto o el porcentaje de programa asimilado por un estudiante en un curso. Las técnicas de control de calidad tienen por objeto mejorar el funcionamiento de los procesos para aumentar la calidad de los resultados obtenidos. Se desarrollaron inicialmente para procesos industriales, campo que constituye todavía su aplicación más frecuente y al que nos referiremos preferentemente en esta sección, pero pue-

den aplicarse para cualquier proceso administrativo, comercial o de servicio en toda organización.

13.1.1 Historia del control de calidad

El comienzo de la fabricación en serie a principios del siglo XX produjo la necesidad de crear especificaciones precisas en los elementos fabricados, que no eran necesarias cuando un artesano era responsable de todo el proceso de fabricación. Por otro lado, la fabricación en serie puso de manifiesto que los procesos de fabricación dan lugar a productos que siempre tienen variabilidad. Walter Shewhart, un ingeniero de Bell Laboratories, descubrió en los años veinte que aunque es inevitable una cierta variabilidad en todos los procesos, podemos controlar y reducir esta variabilidad mediante métodos estadísticos. El gráfico de control que Shewhart diseñó ha contribuido desde entonces a mejorar y controlar innumerables procesos industriales.

Durante la Segunda Guerra Mundial los graves problemas de aprovisionamiento del ejército llevaron a Estados Unidos a crear el Statistical Research Group, que, entre otros estudios estadísticos, estableció reglas precisas basadas en la teoría de contraste de hipótesis para la aceptación de suministros. Fruto de su trabajo fueron las primeras tablas para el control de recepción, las tablas Military Standard, que han sido aceptadas después como estándares internacionales por la International Standard Organization (ISO). W. Edwards Deming, un estadístico que estudió con Shewhart, utilizó los principios de control de procesos para establecer una filosofía de dirección empresarial basada en estas ideas, que tuvo una gran repercusión en Japón, donde Deming estuvo enseñando después de la Segunda Guerra Mundial. Deming enfatizó la necesidad de controlar todos los procesos de una organización y aplicar ideas estadísticas para mejorarlo. Su impacto fue decisivo para mejorar la calidad de los productos japoneses en los años sesenta y setenta, que condujeron a un espectacular crecimiento económico en Japón. En su honor, Japón creó el premio Deming, que se concede anualmente a la organización que haya conseguido mayores mejoras de sus procesos mediante la aplicación de las ideas de control estadístico.

El ejemplo de Japón fue seguido por Estados Unidos, donde se produce en los años ochenta un interés masivo por implantar ideas de control de procesos en todos los niveles de una organización. Como fruto de este cambio, el gobierno de Estados Unidos crea en 1988 el premio Nacional de Calidad Malcom Baldridge para premiar a las empresas que muestran mayores avances en la implantación de métodos de mejora de calidad en toda la organización. En este premio se reconoce que las ideas de calidad son útiles para cualquier proceso, en toda organización. En la misma línea, un grupo destacado de empresas europeas crea en 1991 la Fundación Europea para la Gestión de Calidad (EFQM), que concede también su premio a la excelen-

cia y calidad empresarial. Muchos países han creado además premios similares siguiendo esta filosofía general.

La progresiva caída de las barreras comerciales en todo el mundo y los cambios introducidos a finales del siglo XX en la formación e información de los consumidores por las nuevas tecnologías han colocado la mejora de la calidad de los procesos y servicios como uno de los problemas clave de la llamada nueva economía en el siglo XXI.

13.1.2 Clasificación de los sistemas de control

El control de calidad se clasifica en:

a) Control en curso de fabricación (de procesos).
b) Control de recepción y de producto acabado.

El control en curso de fabricación se realiza continuamente durante la fabricación del producto, a intervalos de tiempo fijos, y tiene por objeto vigilar el funcionamiento del sistema en las mejores condiciones posibles y recoger información para mejorarlo.

El control de recepción se aplica a una partida de nuevo producto, sea éste materia prima, materiales, producto semielaborado o acabado, para inspeccionar que se verifican las especificaciones establecidas.

El control de fabricación produce, a la larga, los mayores beneficios: además de la función de inspección (detectar fallos), que comparte con el control de recepción, permite aprender sobre las causas de variabilidad del proceso, aportando datos para mejorarlo. Por esta razón, el control de fabricación es una herramienta imprescindible para la evaluación de acciones encaminadas a prevenir los posibles fallos y a perfeccionar el proceso productivo.

El control de calidad se realiza observando en cada elemento:

1) Una característica de calidad medible (longitud, resistencia, contenido de impurezas, etc.) que se compara con un estándar fijado. Se denomina entonces control por variables.
2) Un atributo o característica cualitativa que el producto posee o no (como el control pasa/no pasa, por piezas defectuosas, etc.). Se denomina entonces control por atributos.
3) El número total de defectos. Se denomina entonces control por número de defectos.

El control por características medibles o por variables es más informativo que por atributos, ya que indica no sólo si un elemento es o no defectuoso, sino, además, la magnitud del defecto: no es lo mismo que un elemento tenga una longitud fuera de tolerancias por micras que por centímetros. En

consecuencia, es mucho más eficaz para identificar las causas de los problemas de calidad, lo que justifica que se utilice especialmente en el control de procesos. Cuando el objetivo del control no es establecer acciones preventivas, sino únicamente verificar las especificaciones —como ocurre en el control de recepción—, el control por atributos y por número de defectos es más rápido y simple de aplicar y, por tanto, más económico.

13.2 Fundamentos del control de procesos

13.2.1 El concepto de proceso bajo control

Todo proceso de fabricación tiene cierta variabilidad que no puede atribuirse a una causa única, siendo el resultado de los efectos combinados de muchas. Llamaremos a las causas responsables de esta variabilidad *causas no asignables*. Entre éstas, citaremos la variabilidad de la materia prima, la precisión de las máquinas y de los instrumentos de medida, la destreza de los operarios, etc. Estas causas no asignables, que pueden clasificarse en personas, procesos, materiales y métodos, hacen que, al repetir el proceso en condiciones aparentemente análogas, se obtengan resultados distintos.

Existen otras causas de variabilidad que, cuando actúan, producen ciertos efectos previsibles y definidos: por ejemplo, un fallo en una máquina produce elementos defectuosos, pero al ajustarla se elimina la causa de variabilidad y los defectos desaparecen. Llamaremos a estas causas *asignables*, para diferenciarlas de las anteriores.

Todo proceso de funcionamiento regular tiene variabilidad debida a ambos tipos de causas. Las causas no asignables están presentes siempre, produciendo una variabilidad homogénea y estable que es predecible al ser constante. Las asignables sólo intervienen en determinados momentos, y producen entonces una variabilidad muy grande. Los defectos debidos a causas no asignables aparecen aleatoriamente y la aparición de un defecto no hace más probable la aparición del siguiente. Por el contrario, los defectos debidos a causas asignables se mantienen hasta que eliminemos la causa que los produce.

Estudiando un proceso de fabricación es posible eliminar sucesivamente las causas asignables de manera que la variabilidad restante sea debida únicamente a causas no asignables. *Diremos entonces que el proceso se encuentra en estado de control*. Ningún proceso se encuentra espontáneamente en estado de control; llevarlo a dicho estado y *mantenerlo* en él es un logro. Éste es el objetivo del control de procesos.

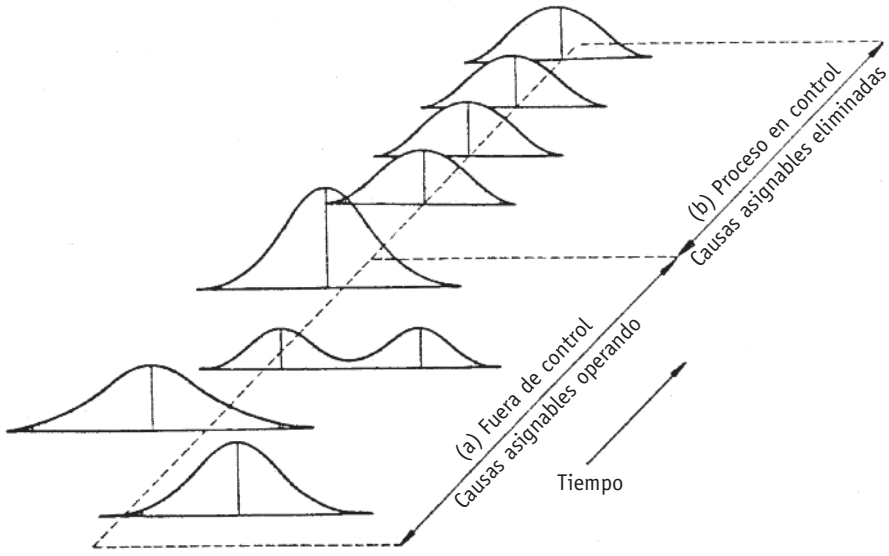
La responsabilidad de eliminar las causas asignables corresponde al supervisor del proceso: el desajuste de una máquina, el error de un operario, etc., son causas directamente detectables y resolubles dentro del marco del proceso productivo. Sin embargo, la responsabilidad de reducir la variabili-

dad producida por las causas no asignables, que son la mayoría, corresponde a la dirección de la empresa: mejorando la tecnología, cambiando los proveedores y, en general, mejorando el proceso productivo.

En resumen, cuando un proceso está en estado de control, la variabilidad es constante a lo largo del tiempo y, por tanto, *predecible*. La proporción de elementos defectuosos es constante a largo plazo y no tiende a aumentar ni a decrecer (véase la figura 13.1b). Por el contrario, cuando el proceso está fuera de control la variabilidad no es constante, siendo sus valores futuros impredecibles (figura 13.1a).

Causas no asignables	Causas asignables
—Existen muchas, cada una de pequeña importancia. —Producen una variabilidad estable. —Es difícil reducir sus efectos.	—Existe un número pequeño pero que produce fuertes efectos. —Producen una variabilidad imprevisible. —Sus efectos desaparecen al eliminar la causa.
Ejemplos:	
Variaciones debidas a la materia prima, a diferencias de habilidad entre los operarios, a factores ambientales.	Variabilidad debida a desajuste, errores humanos, lotes defectuosos, fallos de controles.

Figura 13.1 Procesos bajo control (caso b) y fuera de control (caso a)



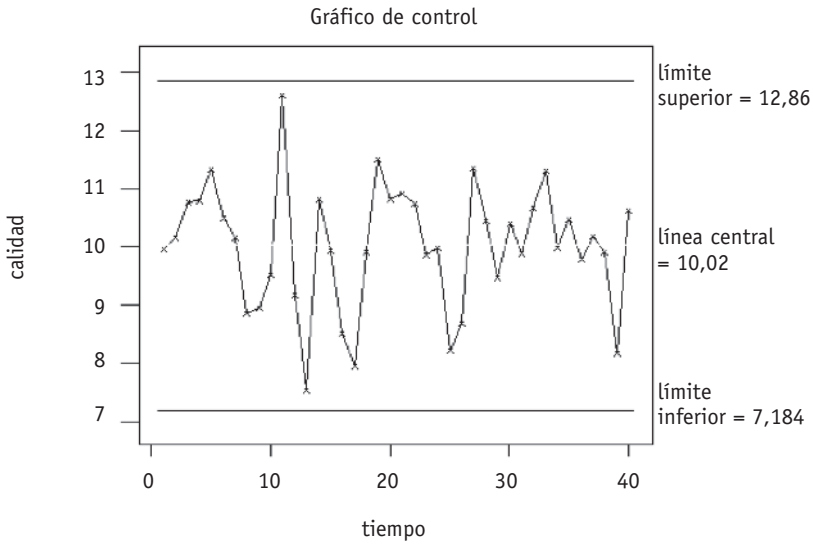
13.2.2 Gráficos de control

La herramienta principal para comprobar si un proceso está en estado de control es el gráfico de control, que controla la evolución de la característica de calidad del proceso a lo largo del tiempo. El gráfico se construye estableciendo una línea central, que representa el valor esperado de la característica de calidad que va a controlarse, y dos líneas laterales, que indican la variabilidad máxima esperada de esta característica de calidad cuando el proceso está en estado de control. Normalmente las líneas laterales se construyen de manera que cuando el proceso esté en estado de control la probabilidad de que una observación salga fuera del intervalo formado por las líneas laterales sea muy baja, del orden del tres por mil.

La figura 13.2 presenta un gráfico de control para una característica de calidad que sigue una distribución normal. Este gráfico puede verse como una herramienta muy efectiva para contrastar en cada muestra que el proceso está en estado de control. Mientras las observaciones se encuentren entre los límites, no hay evidencia para suponer que se ha producido un cambio en el proceso. Por el contrario, cuando se observe un valor fuera de los límites, concluiremos que el proceso está fuera de control y trataremos de descubrir la causa asignable responsable del cambio para eliminarla.

Un gráfico de control puede construirse para observaciones individuales o para promedios de valores. Siempre que sea posible se utilizan más estos últimos, ya que, como veremos, son más eficaces para detectar cambios en el proceso. Los gráficos de control más utilizados se exponen a continuación.

Figura 13.2 Gráfico de control



13.3 El control de procesos por variables

13.3.1 Introducción

Supondremos en esta sección que un proceso produce elementos cuya calidad está descrita por una característica medible x. Por ejemplo, x puede ser en un proceso de mantenimiento el tiempo transcurrido hasta completar el servicio; en un proceso de fabricación, la longitud de una pieza, la resistencia de un circuito o la capacidad de un chip; en un proceso de servicio, el tiempo de servicio o el grado de satisfacción de los usuarios medido por una encuesta; en un proceso docente, el aprendizaje adquirido por los estudiantes de acuerdo con cierta escala de medida.

Supondremos que el proceso está diseñado para proporcionar, por término medio, una característica de calidad que llamaremos valor nominal o de diseño y que representaremos por μ. Por ejemplo, si x es un tiempo de servicio, μ será el tiempo medio fijado para ese servicio. En general podemos suponer que el proceso (máquinas, herramientas, personal, controles, etc.) se diseña o se ajusta de manera que la distribución de x en la fabricación está centrado en μ, valor nominal.

Todos los procesos tienen variabilidad y por tanto los resultados del proceso no serán siempre idénticos. El primer paso para controlar el proceso es estimar esta variabilidad. A continuación construiremos gráficos de control para comprobar si los resultados del proceso están centrados en el valor nominal y si la variabilidad permanece constante. En estas condiciones, si suponemos que la distribución de los resultados es conocida, puede calcularse a priori la proporción de la fabricación que estará entre dos límites fijos. Cuando esto ocurre, el proceso está en estado de control y sus resultados son predecibles.

13.3.2 Determinación de la variabilidad del proceso

Determinar la variabilidad del proceso requiere estimar la desviación típica de la distribución de su característica de calidad. Para ello, se toman observaciones fabricadas en *condiciones normales de operación*, tratando de eliminar las causas asignables de variación, de manera que las muestras correspondan a un proceso en condiciones de control estadístico. Los datos deben tomarse durante un tiempo suficientemente dilatado, para incluir todas las posibles causas esperables de variación: cambios de turnos, fatiga de los operarios, distintos proveedores de materia prima, etc.

Como el proceso puede pasar inadvertidamente a una situación fuera de control (por ejemplo, por desajuste de herramientas), es conveniente tomar varias muestras pequeñas igualmente espaciadas a lo largo del intervalo de producción (cada hora, 2 horas, día, etc.). Los elementos de cada muestra

se toman consecutivos, para que sean lo más homogéneos posible. Llamaremos x_{ij} al valor de la característica de calidad en el elemento j de la muestra i, y supondremos que tenemos k muestras, cada una de n elementos:

$$(x_{11}, ..., x_{1n}), (x_{21}, ..., x_{2n}), (x_{k1}, ..., x_{kn})$$

Si el proceso hubiera permanecido bajo el control durante todo el período de recogida de información, estos nk datos constituirían una muestra aleatoria simple de la misma población. El valor medio de la característica de calidad se estimará por la media de estos nk datos:

$$\overline{\overline{x}} = \frac{\Sigma\Sigma\, x_{ij}}{nk}$$

y la variabilidad mediante la varianza:

$$\hat{s}^2 = \frac{\Sigma\Sigma\, (x_{ij} - \overline{\overline{x}})^2}{nk - 1}$$

Sin embargo, recordemos que el estado de control estadístico es un logro, un objetivo a alcanzar, y no el estado natural del proceso. En consecuencia, es probable que las k muestras no provengan de la misma población, ya que durante el intervalo de recogida de información el proceso puede haber pasado a una situación de falta de control, por cambios en la medida o en la variabilidad. Para decidir respecto a este aspecto, se utilizan los dos gráficos de control que describimos a continuación.

13.4 Gráficos de control por variables

13.4.1 Gráfico de control para medias

El gráfico de control para las medias se utiliza para comprobar si un conjunto de muestras del proceso provienen de una distribución con la misma media. Diremos entonces que las observaciones son homogéneas en la media. Supongamos que tenemos k muestras de tamaño n de un proceso. El gráfico de medias se calcula como sigue:

1. *Calcular la media y desviación típica de cada muestra*. Sean éstas $(\overline{x}_1, ..., \overline{x}_k)$; $(s_1, ..., s_k)$, donde, por ejemplo:

$$\overline{x}_1 = \frac{\Sigma x_{1j}}{n} \qquad s_1^2 = \frac{\Sigma (x_{1j} - \overline{x}_1)^2}{n}$$

2. *Estimar la media y desviación típica del proceso suponiendo homogeneidad*. Si todas las observaciones provienen de la misma distribución, la media de la característica de calidad en el proceso se estima por:

$$\bar{\bar{x}} = \frac{\Sigma \bar{x}_i}{k} = \frac{\Sigma\Sigma\, x_{ij}}{N} \tag{13.1}$$

donde $N = kn$. Este estimador será centrado si el proceso tiene una media μ constante.

Para estimar la desviación típica del proceso, σ, tendremos en cuenta que s_i no es un estimador centrado de σ, y que tiende a subestimar la variabilidad del proceso. En efecto, la variabilidad de una muestra pequeña será, en promedio, menor que la variabilidad existente en la población. El estimador s_i tendrá un sesgo de subestimación tanto mayor cuanto menor sea el tamaño muestral n. Se demuestra en el apéndice 13A que:

$$E\,[s_i] = c_2\sigma$$

donde los coeficientes c_2, que son menores que la unidad, se encuentran tabulados en función del tamaño muestral en la tabla 13.1. Por ejemplo, si $n = 4$ se obtiene que $c_2 = 0{,}7979$, lo que indica que en promedio la desviación típica de una muestra de tamaño 4 es sólo el 79,79% de la desviación típica en la población. Por lo tanto, para estimar σ tendremos que corregir el estimador s mediante s_i/c_2, que será un estimador centrado de σ. Por ejemplo, con $n = 4$, obtendremos que $1.2533 s_i$ es un estimador centrado. Si suponemos que las poblaciones que han generado las muestras tienen la misma variabilidad, podemos estimar esta variabilidad común promediando los estimadores centrados que obtenemos con cada muestra. El estimador resultante será:

$$\hat{\sigma} = \frac{\Sigma\, s_i/c_2}{k} = \bar{s}/c_2 \tag{13.2}$$

donde $\bar{s} = \Sigma\, s_i/k$ es el promedio de las desviaciones típicas. El estimador $\hat{\sigma}$ es un estimador centrado de σ que utiliza toda la información disponible.

3. *Contrastar si todas las medias son homogéneas*. Cuando todas las muestras provienen de la misma población, $(\bar{x}_j - \mu)\ \sqrt{n}/\sigma$ sigue una distribución normal estándar. Si el número de datos totales $N = nk$ es grande, digamos mayor que 100, al sustituir en la expresión anterior los parámetros (μ, σ) por sus estimaciones, $(\bar{\bar{x}}\hat{\sigma})$ obtendre-

mos también aproximadamente una distribución normal estándar, es decir:

$$(\bar{x}_j - \bar{\bar{x}})\ \sqrt{n}/\hat{\sigma} \sim N(0, 1).$$

Por tanto, como con el 99% de probabilidad una variable normal no debe alejarse de su media más de tres desviaciones típicas, podemos prever que si el proceso está en estado de control las medias muestrales con el 99% de probabilidad deben estar en el intervalo:

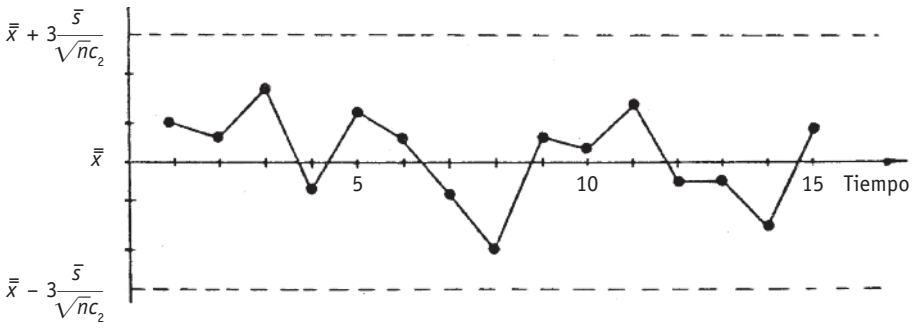
$$\bar{x}_j \in \bar{\bar{x}} \pm 3\hat{\sigma}/\sqrt{n}$$

El contraste de que las k medias provienen de la misma población y, por tanto, el proceso está bajo control se realiza comprobando que todas las medias $\bar{x}_i$ están incluidas en el intervalo $\bar{\bar{x}} \pm 3\hat{\sigma}/\sqrt{n}$. Para ello, se construye el gráfico de control de la media, cuya línea central es $\bar{\bar{x}}$ y cuyas líneas laterales estarán situadas simétricamente respecto a la central a una distancia de $3\hat{\sigma}/\sqrt{n}$. A continuación llevaremos a este gráfico en abcisas el tiempo, o el número de muestra, y en ordenadas los valores $\bar{x}_j$ (véase la figura 13.3). Si alguna media muestral sale fuera de los límites, concluiremos que esa muestra no es homogénea con las anteriores y que en el momento en que se ha tomado el proceso estaba fuera de control.

Para clarificar el procedimiento anterior conviene hacer dos comentarios:

(1) Si en lugar de calcular las desviaciones típicas muestrales con s_i hubiésemos utilizado las desviaciones típicas corregidas por grados de li-

Figura 13.3 Gráfico para control de la media



bertad, $\hat{s}_i = \sqrt{\Sigma(x_{1j} - \bar{x}_1)^2/(n-1)}$, el procedimiento para estimar la desviación típica de la población sería análogo, pero habría que modificar la constante c que corrige el sesgo de subestimación, ya que el sesgo de $\hat{s}_i$ es menor que el de s_i. La nueva corrección resulta inmediatamente de la relación entre ambos estimadores. Como $\hat{s}_i = s_i \sqrt{n}/\sqrt{n-1}$, si s_i/c_2 es un estimador centrado también lo será $\hat{s}_i \sqrt{n-1}/\sqrt{n}c_2$. Por ejemplo, con muestras de tamaño 4, obtenemos un estimador centrado con $\hat{s}_i\sqrt{3}/2(0{,}7979) = 1.0854\hat{s}_i$ que será igual numéricamente a $s_i(0{,}7979)$. En consecuencia la estimación de $\hat{\sigma}$ se calcularía mediante $\Sigma\hat{s}\sqrt{n-1}_j/c_2k\sqrt{n}$.

(2) Si la varianza de la población hubiese permanecido constante durante todo el período en que se han tomado las muestras, podría pensarse que en lugar del estimador $\bar{s}$ sería mejor utilizar como estimador (con su corrección por sesgo) $\hat{s}_T$ dado por:

$$\hat{s}_T = \sqrt{\frac{\Sigma\Sigma(x_{ij} - \bar{x}_j)^2}{N-k}} = \sqrt{\frac{n}{N-k}\sum s_j^2} = \sqrt{\frac{1}{k}\sum \hat{s}_j^2}$$

ya que $\hat{s}_T^2$ es un buen estimador (centrado y con alta precisión) para σ^2. Sin embargo, si comparamos ambos estimadores vemos que $\bar{s}$ es un promedio de desviaciones típicas y $\hat{s}_T$ es la raíz cuadrada de un promedio de varianzas. Cuando alguna de las observaciones es heterogénea con las demás, por ejemplo, mucho mayor que el resto, el primer estimador se ve menos afectado (es más robusto) que el segundo. Por ejemplo, con $n = 5$, las desviaciones típicas, s_i (10, 11, 9, 10, 60) que incluyen un valor anormal, 60, conducen a:

$$\hat{\sigma} = \bar{s}/c_2 = 20/0{,}84 = 23{,}4; \qquad \hat{s}_T = 31{,}6$$

y $\hat{s}_T$ está bastante más distorsionada que $\hat{\sigma}$ por el valor extremo.

13.4.2 Gráfico de control para desviaciones típicas

El control de la variabilidad se realiza estudiando o bien la desviación típica, o bien el rango de la muestra. Expondremos aquí el gráfico de la desviación típica y, en la sección siguiente, el del rango.

La variabilidad en cada muestra puede medirse por s_i, desviación muestral sin corregir por grados de libertad, o por $\hat{s}_i$, la corregida por grados de libertad. En ambos casos el análisis es muy similar. Suponiendo que utilizamos s_j y llamando como en la sección anterior $\bar{s} = \Sigma s_i/k$, puede demostrarse (apéndice 13A) que un intervalo aproximado del 99% para estas desviaciones es:

$$(B_3\bar{s}, B_4\bar{s})$$

Tabla 13.1 Factores para calcular líneas de gráficas de control utilizando la desviación típica muestral sin corregir

Número de observaciones en muestra, *n*	Gráfico para desviaciones estándares			Gráfico para rangos				
	Factor para línea central	Factores para límites de control		Factor para línea central	Factores para límites de control			
	C_2	B_3	B_4	d_2	D_1	D_2	D_3	D_4
2	0,5642	0,000	3,267	1,128	0,000	3,686	0,000	3,276
3	0,7236	0,000	2,568	1,693	0,000	4,358	0,000	2,575
4	0,7979	0,000	2,266	2,059	0,000	4,698	0,000	2,282
5	0,8407	0,000	2,089	2,326	0,000	4,918	0,000	2,115
6	0,8686	0,030	1,970	2,534	0,000	5,078	0,000	2,004
7	0,8882	0,118	1,882	2,704	0,205	5,203	0,076	1,924
8	0,9027	0,185	1,815	2,847	0,387	5,307	0,136	1,864
9	0,9139	0,239	1,761	2,970	0,546	5,394	0,184	1,816
10	0,9227	0,284	1,716	3,078	0,687	5,469	0,223	1,777
11	0,9300	0,321	1,679	3,173	0,812	5,534	0,256	1,744
12	0,9359	0,354	1,646	3,258	0,924	5,592	0,284	1,719
13	0,9410	0,382	1,618	3,336	1,026	5,646	0,308	1,692
14	0,9453	0,406	1,594	3,407	1,121	5,693	0,329	1,671
15	0,9490	0,428	1,572	3,472	1,207	5,737	0,348	1,652
16	0,9523	0,448	1,552	3,532	1,285	5,779	0,364	1,636
17	0,9551	0,466	1,534	3,588	1,359	5,817	0,379	1,621
18	0,9576	0,482	1,518	3,640	1,426	5,854	0,392	1,608
19	0,9599	0,497	1,503	3,689	1,490	5,888	0,404	1,596
20	0,9619	0,510	1,490	3,735	1,548	5,922	0,414	1,586
21	0,9638	0,523	1,477	3,778	1,606	5,950	0,425	1,575
22	0,9655	0,534	1,466	3,819	1,659	5,979	0,434	1,566
23	0,9670	0,545	1,455	3,858	1,710	6,006	0,443	1,557
24	0,9684	0,555	1,445	3,895	1,759	6,031	0,452	1,548
25	0,9696	0,565	1,435	3,931	1,804	6,058	0,459	1,541

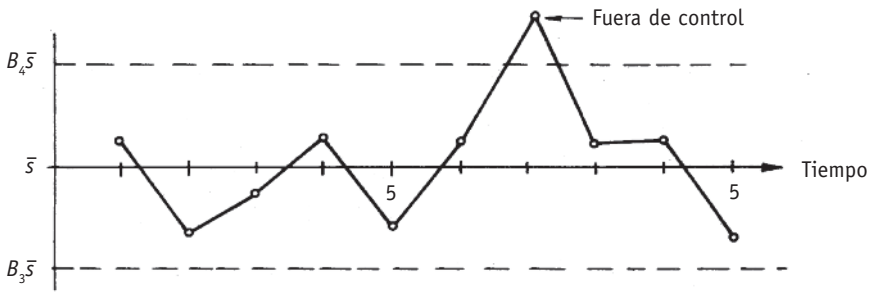
donde los coeficientes B_3 y B_4 se obtienen de la tabla 13.1. Además, el valor esperado de s_j, supuesto que todas las muestras provienen de poblaciones normales con la misma varianza (aunque quizás distinta media), se estimará por $\overline{s}$.

El procedimiento operativo para construir el gráfico es el siguiente:

1. Dado n y el valor $\overline{s}$ calculado para el gráfico de la media, obtener de la tabla 13.1 los valores B_3 y B_4. Calcular los límites superior, $B_4\overline{s}$, e inferior $B_3\overline{s}$.
2. Construir el gráfico representando en abscisas el tiempo y en ordenadas los valores s_i. Marcar el gráfico con las líneas central ($\overline{s}$) y de control ($B_3\overline{s}$, $B_4\overline{s}$) y representar las desviaciones típicas de las muestras, s_i (figura 13.4).
3. Si alguna de las desviaciones sale fuera de los límites de control, admitir que dicha muestra no es consistente con las demás.

En el apéndice 13A se justifica este gráfico y se demuestra que utilizando las desviaciones $\hat{s}_i$, el gráfico tiene por línea central $\hat{\overline{s}} = \Sigma\hat{s}_i/k$ y por límites ($B_3\hat{\overline{s}}$, $B_4\hat{\overline{s}}$).

Figura 13.4 Gráfico para la desviación típica



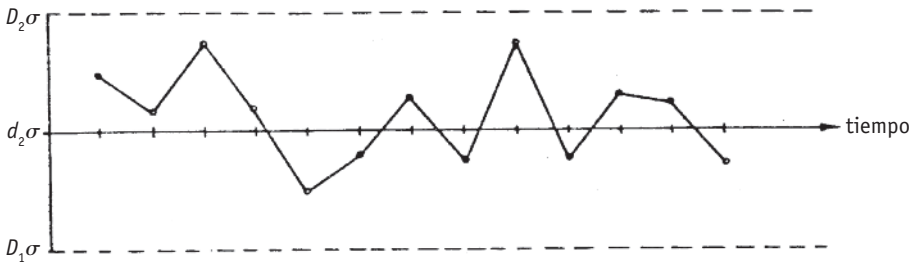
13.4.3 Gráfico de control para rangos

El control de la variabilidad puede hacerse mediante la desviación típica, como hemos presentado anteriormente, pero es mucho más frecuente que se realice utilizando el rango de la muestra. La razón es que en muestras pequeñas el rango es casi tan eficiente como la desviación típica, siendo mucho más fácil de calcular. Cuando las muestras se tomen y analicen de forma automática con un ordenador, o cuando el tamaño muestral sea mayor de cinco o seis unidades, deben utilizarse los gráficos de la desviación típica antes expuestos. En cualquier otro caso, los gráficos del rango tienen la ventaja de su simplicidad.

Recordemos que el rango de una muestra es la diferencia entre el valor mayor y el menor. Al tomar muestras de una población normal, el rango muestral sigue una distribución que puede calcularse y que tiene media $d_2\sigma$, siendo d_2 una constante que depende del tamaño muestral y que está tabulada. También están tabulados en función de n los coeficientes D_1, D_2 (tabla 13.2), que definen un intervalo en el que se debe encontrar el rango muestral con probabilidad 99%.

En el caso, no muy frecuente, en que la desviación típica del proceso, σ, es conocida (por ejemplo, por estudios anteriores), el control de la variabilidad mediante el rango utiliza el gráfico de control de la figura 13.5. Cuando σ sea desconocido, se estima a partir de los rangos muestrales como sigue.

Figura 13.5 Gráfico de rangos



Llamando

$$\overline{\mathrm{R}} = \frac{\Sigma R_i}{k}$$

al rango medio de todas las muestras, la desviación típica del proceso se estima por $\overline{\mathrm{R}}/d_2$, donde los coeficientes d_2 se encuentran en la tabla 13.1. Sustituyendo $\overline{\mathrm{R}}/d_2$ por σ en los gráficos de control de la media, éstos tendrán como límites:

$$\bar{x} \pm 3\frac{\hat{\sigma}}{\sqrt{n}} = \bar{x} \pm \frac{3}{\sqrt{n}}\frac{\overline{\mathrm{R}}}{d_2}$$

Análogamente, el gráfico de variabilidad mediante rangos tiene una línea central $\overline{\mathrm{R}}$ y líneas de control $D_3\overline{\mathrm{R}}$, $D_4\overline{\mathrm{R}}$. Todos estos coeficientes están en la tabla 13.1.

Estos resultados se resumen en la tabla 13.2.

Tabla 13.2 Fórmulas para líneas centrales y límites de control

Gráfico de	Variabilidad media por	Línea central	Límites
Medias	Desviaciones típicas	$\overline{x}$	$\overline{x} \pm 3\overline{s}/\sqrt{n}c_2$
Medias	Rangos	$\overline{x}$	$\overline{x} \pm 3\overline{\mathbf{R}}/(d_2\sqrt{n})$
Desviaciones típicas	Desviaciones típicas	$\overline{s}$	$B_3\overline{s},\ B_4\overline{s}$
Rango	Rango	$\overline{\mathbf{R}}$	$D_3\overline{\mathbf{R}},\ D_4\overline{\mathbf{R}}$

13.4.4 Estimación de las características del proceso

Si alguna muestra aparece fuera de los límites de control en cualquiera de los dos gráficos, la eliminaremos, ya que indica que el proceso, en dicho instante, estaba fuera de control. A continuación recalcularemos $\overline{x}$ y $\overline{s}$ con las muestras restantes, construiremos nuevos gráficos y comprobaremos si ahora todas las muestras son aparentemente homogéneas. Si no lo son, eliminaremos las heterogéneas, y repetiremos los cálculos hasta obtener un grupo homogéneo.

Con las muestras finales obtendremos una estimación inicial de la media y la variabilidad del proceso. La estimación de σ será $\overline{s}/c_2$, donde $\overline{s}$ incluye sólo las muestras que están dentro de los límites de control y los coeficientes c_2 están tabulados (tabla 13.1).

A continuación contrastaremos la normalidad de la distribución de la variable. Este contraste puede efectuarse con cualquiera de los tests estudiados anteriormente. El test más utilizado en control de procesos es dibujar los puntos en papel probabilístico normal y comprobar si siguen una recta. También puede utilizarse cualquiera de los contrastes de bondad de ajuste estudiados anteriormente. Si la distribución de los datos no es normal, esto puede indicar que el proceso es muy heterogéneo, estando sometido a causas asignables que deberíamos identificar y controlar. De todas formas, si no se dispone de más información, podemos iniciar el control del proceso mediante los gráficos de control e ir mejorando paulatinamente su funcionamiento. En estos casos es esperable que el estimador de la desviación típica obtenido por el método expuesto será poco preciso, y utilizaremos esta estimación provisionalmente para comenzar a controlar el proceso, según estudiaremos en la sección siguiente. Después de cierto tiempo, este control permite identificar y eliminar los efectos de las causas asignables, con lo que podemos repetir el estudio para calcular una estimación más exacta de la variabilidad del proceso.

Ejemplo 13.1

Para determinar la variabilidad de un proceso se toman 25 muestras de tamaño 6 y se calculan la media y la desviación típica en cada una de ellas con los resultados siguientes (los resultados se dan redondeados para facilitar los cálculos).

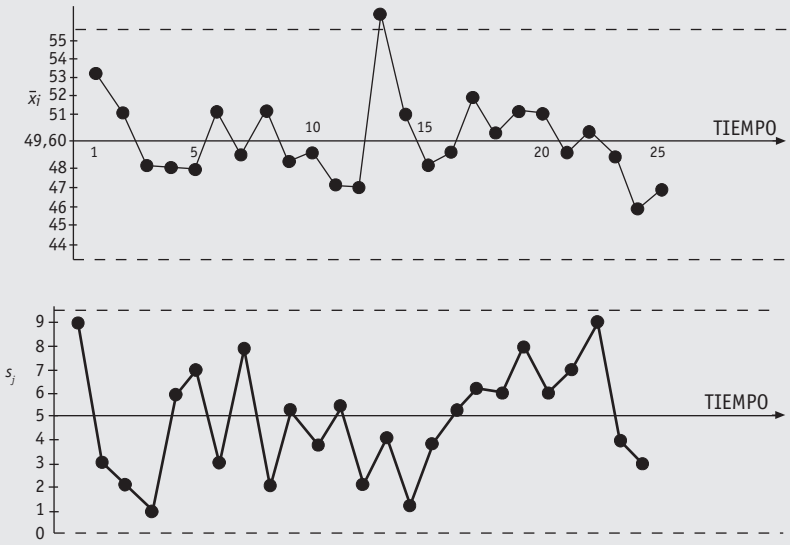
N.º de muestra	1	2	3	4	5	6	7	8	9	10	11	12	13
$\bar{x}$	53	51	48	48	48	41	49	51	48	49	47	47	57
s_j	9	3	2	1	6	7	3	8	2	5	4	5	2
N.º de muestra	14	15	16	17	18	19	21	21	22	23	24	25	
$\bar{x}$	51	48	49	52	50	51	51	49	50	49	46	47	
s_j	4	1	4	5	6	6	8	6	7	9	4	3	

$$\bar{\bar{x}} = 49{,}60 \; ; \; \bar{s} = 4{,}8 \; ; \; \hat{\sigma} = 4{,}8/0{,}8686 = 5{,}53$$

Los gráficos de control serán:

a) Media: $49{,}60 \pm 3 \cdot 5{,}53/\sqrt{6} = 49{,}60 \pm 6{,}77$, es decir (42,83; 56,37).
b) Desviación: $B_3 = 0{,}03$; $B_4 = 1{,}97$; por tanto, $B_3\bar{s} = 0{,}14$; $B_4\bar{s} = 9{,}6$.

Figura 13.6 Gráficos de control para el ejemplo 13.1



La muestra n.º 13 tiene una media que sale fuera de los límites de control (figura 13.6). La desviación típica de dicha muestra es normal, por lo que sospechamos un desplazamiento de la media. Eliminando dicha muestra, las 24 resultantes conducen a:

$$\bar{\bar{x}} = 49{,}29 \; ; \; \bar{s} = 4{,}92 \; ; \; \hat{\sigma} = 4{,}92/0{,}8686 = 5{,}66$$

y a los nuevos límites:

a) Media: $49{,}29 \pm 3 \cdot 5{,}66/\sqrt{6}$; intervalo (42,36 ; 56,22).
b) Desviación: (0,15 ; 9,69).

Todos los puntos se encuentran ahora dentro de control. La estimación de la desviación típica del proceso es, por tanto:

$$\hat{\sigma} = 4{,}92/0{,}8686 = 5{,}66$$

Sería conveniente construir un histograma de los datos originales para contrastar la normalidad de la fabricación.

13.5 Implantación del control por variables

Una vez determinadas las características del proceso, se comienza el control de la fabricación. Este control puede verse como un contraste continuado de la hipótesis de que el proceso está en estado de control. Este contraste se aplica a cada muestra que se toma a intervalos regulares de tiempo. Si la hipótesis es cierta, y el proceso está bajo control, la media y la desviación típica muestrales tienen que estar incluidas dentro de los límites de control de sus gráficos respectivos. Como estos gráficos se calculan con tres desviaciones típicas, la probabilidad de cometer un error tipo I y rechazar que el proceso está en estado de control cuando realmente lo está es del orden del uno por mil en cada contraste. También podemos calcular la probabilidad de que este gráfico sea capaz de detectar desajustes en el proceso en función del tamaño de estos desajustes, como veremos a continuación. La frecuencia de muestreo depende de la relación entre la variabilidad del proceso y las tolerancias del producto, como veremos posteriormente.

Cuando las medidas de una muestra salen fuera de los límites de control, rechazaremos la hipótesis de estado de control e investigaremos las causas de este hecho para evitarlas en el futuro. Vamos a estudiar el funcionamiento de estos gráficos.

13.5.1 Eficacia del gráfico de la media

El control de procesos se realiza siempre que es posible mediante medias muestrales porque éstas son más eficaces para detectar cambios que las observaciones individuales. En efecto, supongamos que la media de un proceso se desajusta dos desviaciones típicas, es decir, el proceso pasa de fabricar según una distribución normal de media μ_0 y desviación σ a hacerlo mediante una normal con media $\mu_1 = \mu_0 + 2\sigma$ y la misma desviación típica. Esta situación podemos representarla gráficamente como indica la figura 13.7. Vamos a comparar la eficacia de los gráficos de observaciones individuales y el de la media para detectar este cambio.

Comencemos con el gráfico de observaciones individuales. En este caso el gráfico tiene línea central μ_0 y límites de control en $\mu_0 \pm 3\sigma$. La probabilidad de que un elemento producido cuando el proceso está desajustado salga fuera del intervalo $\mu_0 \pm 3\sigma$ será, aproximadamente, del 16% (área a la derecha de $\mu_0 + 3\sigma$ en la figura 13.7). Ésta es la probabilidad de que una variable $N(\mu_0 + 2\sigma, \sigma^2)$ sea mayor que $\mu_0 + 3\sigma$, que es equivalente a la probabilidad de que una variable normal estándar sea mayor que uno. Supongamos que observamos cuatro observaciones consecutivas del proceso desajustado y que las representamos en este gráfico. La probabilidad de que las cuatro estén dentro del intervalo de control será $(1 - 0{,}16)^4 = 0{,}48$, y la probabilidad de que una o más de las cuatro observaciones salga fuera de los límites será $1 - 0{,}48 = 0{,}52$, que es la probabilidad de detectar un cambio en la media con cuatro observaciones utilizando el gráfico de observaciones individuales.

En el gráfico de medias de tamaño cuatro representamos la media de las cuatro observaciones, $\bar{x}$, en lugar de los valores individuales. Como la desviación típica de la media muestral es $\sigma/2$, los límites de 3 desviaciones típicas para las medias muestrales serán $\mu_0 \pm 3(\frac{\sigma}{2})$. Cuando la media del proceso se desplace 2σ, la distribución de las medias se desplazará esta misma cantidad; pero ahora, al tener menor desviación, la mayor parte de ella quedará fuera de los límites de control, como indica la figura 13.8. La probabilidad de detectar un cambio de magnitud 2σ con la media muestral de cuatro observaciones será igual a la probabilidad de que la media muestral salga de los límites de control, es decir, la probabilidad de que una variable $N(\mu_0 + 2\sigma, \sigma^2/4)$ sea mayor que $\mu_0 + 3\sigma/2$, que es equivalente a la probabilidad de que una variable normal estándar sea mayor que menos uno, y esta probabilidad es 0,84.

En consecuencia, utilizando la misma información, con el gráfico de observaciones individuales tenemos una probabilidad de detectar este cambio de

0,52, mientras que con el gráfico de medias esta probabilidad aumenta a 0,84. Podemos concluir que se utiliza más eficazmente la información al controlar por medias que por observaciones individuales. El tamaño muestral que se considera en el control por medias es, generalmente, entre 4 y 8 unidades. La práctica ha demostrado que estos tamaños muestrales combinan la rapidez y facilidad de recogida con una razonable sensibilidad para detectar cambios.

Una ventaja adicional de controlar la media del proceso mediante medias muestrales es que, por el teorema central del límite, la distribución muestral de la media será aproximadamente normal, sea cual sea la distribución de las observaciones individuales. Esto justifica la amplia utilización de los gráficos de medias de la figura 13.3. El control de las observaciones individuales requiere, sin embargo, el conocimiento de la distribución de la población.

Figura 13.7 Efecto de un desplazamiento de 2σ en la media sobre la distribución de las observaciones individuales

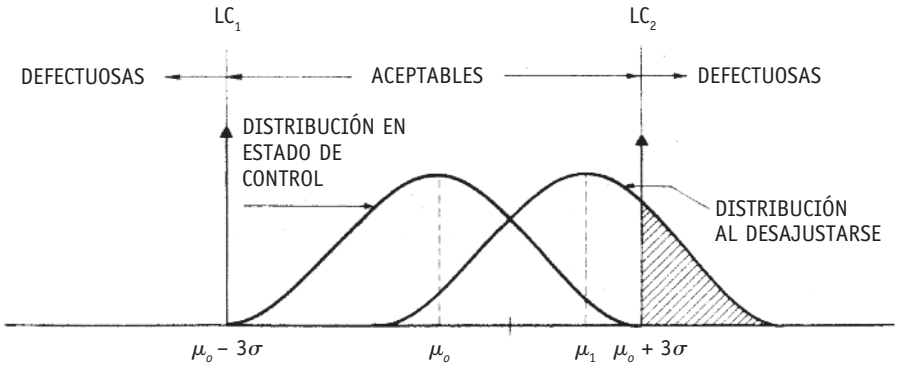
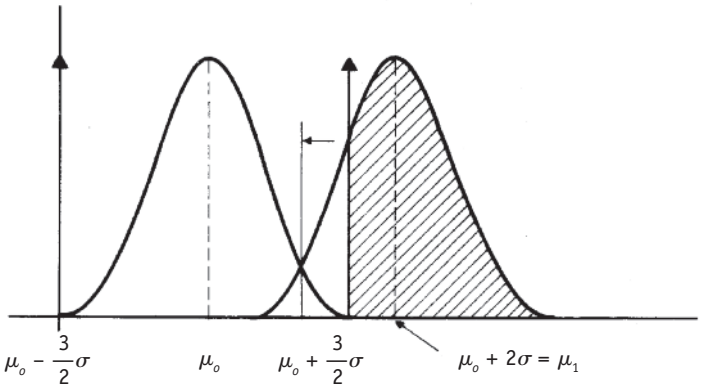


Figura 13.8 Efecto de un desplazamiento 2σ en la distribución de la media muestral



Ejemplo 13.2

Un proceso fabrica en condiciones de control con media 100 mm y desviación típica 5 mm. Calcular la probabilidad de detectar un cambio en la media de 10 mm (2σ) si:

1) El control se realiza con observaciones individuales.
2) El control se realiza con medias de cuatro observaciones.

La proporción de elementos fuera del intervalo de control (85,115) será la porporción del área que una normal $N(110; 5)$ deja fuera de dicho intervalo. Calculando el área dentro:

$$z_a = \frac{85 - 110}{5} = -5 \qquad z_b = \frac{115 - 110}{5} = 1$$

Área hasta 1	0,8413
Área hasta –5	0,0001
	0,8412

la diferencia hasta uno será el área fuera del intervalo de control. El 16% de los elementos saldrán fuera del intervalo $\pm 3\sigma$ construido bajo la hipótesis de que el proceso está bajo control; tendremos que observar por término medio 6 unidades (100/16) para que aparezca un valor fuera de dicho intervalo.

Para las medias, los límites de tres desviaciones típicas son:

$$100 \pm 3 \cdot \frac{5}{\sqrt{n}}$$

Para muestras de tamaño 4 estos límites son (92,5; 107,5). Si el proceso se desajusta pasando a una media 110, las medias muestrales vendrán de una distribución $N(110, 2,5)$. Los límites anteriores equivalen ahora a:

$$z_a = \frac{92,5 - 100}{2,5} = -7 \quad ; \quad z_b = \frac{107,5 - 110}{2,5} = -1$$

El área hasta –7 es prácticamente cero, y el área hasta –1 es 0,16, por lo que habrá una probabilidad del 84% de detectar el cambio con una sola muestra de tamaño 4.

Si repetimos las operaciones con muestras de tamaño 8 y 116, llegamos a la tabla siguiente:

	Probabilidad de detección de un cambio de 2σ en la media del proceso	
Número de observaciones	al considerarlas aisladas	al considerar su media
4	0,52	0,84
8	0,77	0,996
16	0,95	1

13.5.2 Curva característica de operación

Al utilizar un gráfico de control de procesos es importante conocer qué situaciones de fuera de control podemos ser capaces de identificar con rapidez y cuáles van a requerir muchas muestras para ser identificadas. La sensibilidad de un gráfico de control se establece mediante la curva OC, o curva característica de operación, que describe la probabilidad de que una muestra esté dentro de los límites de control para cada posible situación del proceso. Vamos a ilustrar el cálculo de esta curva para el gráfico de medias.

Por construcción, la probabilidad de que la media muestral esté dentro de los límites de control cuando la media verdadera es μ es 0,9913. Vamos a calcular la probabilidad de que la media muestral esté dentro de los límites ante distintos desplazamientos de la media del proceso, pero manteniendo la variabilidad constante. Supondremos que el tamaño muestral es 4, de manera que la desviación típica de la media muestral es $\sigma/\sqrt{4} = 0{,}5\sigma$ y los límites de control están a una distancia de μ de $3\sigma/\sqrt{4} = 1{,}5\sigma$.

Supongamos un desplazamiento de la media de $0{,}5\sigma$. La media del proceso pasa a ser $\mu + 0{,}5\sigma$, pero la desviación típica no varía. La probabilidad de que la media muestral esté entre los límites es igual a la probabilidad de que no sobrepase el límite superior de $1{,}5\sigma$. Denotemos por $P[x \leq a|N(\mu;\sigma)]$ a la probabilidad de que una variable normal con distribución $N(\mu;\sigma)$ sea menor que a. Podemos escribir

$$P[\bar{x} \leq \mu + 1{,}5\sigma|N(\mu + 0{,}5\sigma;\ 0{,}5\sigma) = P(z \leq 1/0{,}5|N(0,1)] = P(z \leq 2) = ,97725$$

es decir, la probabilidad de que la observación caiga dentro de los límites de control y no detectemos el cambio es aproximadamente 0,98, o, en otros términos, la probabilidad de notarlo es 0,02. Por tanto el gráfico de control

es incapaz de detectar este tipo de desplazamientos, ya que necesitaremos en promedio 1/0,02 = 50 muestras para detectar este cambio. Si el desplazamiento es de una desviación típica, tendremos

$$P[\overline{x} \leq \mu + 1{,}5\sigma | N(\mu + \sigma;\ 0{,}5\sigma) = P(z \leq ,1 | N(0,1)] = 0{,}8413$$

y la probabilidad de detectarlo es todavía muy pequeña, 0,1587, lo que supone que necesitaremos en promedio 1/0.1587 = 6,3 muestras para detectarlo. Vemos que un desplazamiento en la media de una desviación típica será detectado en promedio seis períodos después de que ocurra, lo que dificultará su identificación. Para un desplazamiento de $1{,}5\sigma$, tendremos que

$$P[\overline{x} \leq \mu + 1{,}5\sigma | N(\mu + 1{,}5\sigma;\ 0{,}5\sigma) = P(z \leq 0 | N(0,1)] = 0{,}5$$

y la probabilidad de que esté dentro baja a 0,5. Para desplazamientos de dos desviaciones típicas

$$P[\overline{x} \leq \mu + 1{,}5\sigma | N(\mu + 2\sigma;\ 0{,}5\sigma) = P(z \leq -1 | N(0,1)] = 0{,}15813.$$

y serán fácilmente detectables. Estos resultados se resumen en la figura 13.9, que indica la curva OC para estos y otros valores de la media del proceso. Se observa que un gráfico de control para la media con 4 observaciones sólo será capaz de detectar eficientemente desplazamientos en la media superior a dos desviaciones típicas.

Figura 13.9 Curva OC para la media

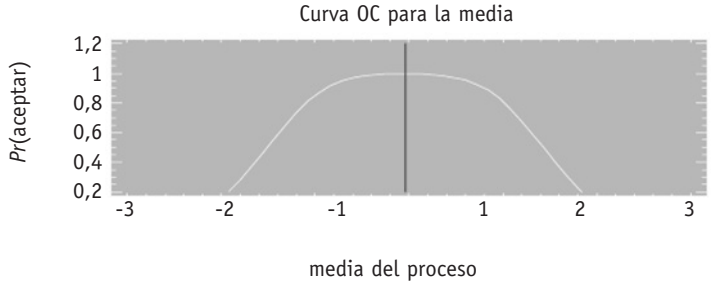


Tabla 13.3 Fórmulas para líneas centrales y límites de control

Gráfico de	Variabilidad media por	Línea central	Límites
Medias	Desviaciones típicas	$\overline{\overline{x}}$	$\overline{\overline{x}} \pm 3\overline{s}/c_2\sqrt{n}$
Medias	Rangos	$\overline{\overline{x}}$	$\overline{\overline{x}} + 3\overline{R}/d_2\sqrt{n}$
Desviaciones típicas	Desviaciones típicas	$\overline{s}$	$B_3\overline{s}$, $B_4\overline{s}$
Rango	Rango	$\overline{R}$	$D_3\overline{R}$, $D_4\overline{R}$

13.5.3 Interpretación de gráficos de control

Los cambios en el funcionamiento del proceso se identificarán en los gráficos de media y variabilidad por pautas específicas que vamos a analizar.

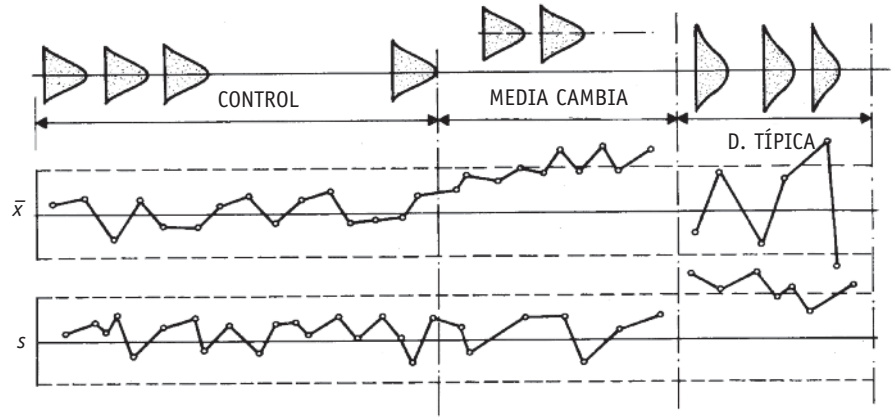
a) Cambios bruscos en la media y/o la variabilidad

Si la media y/o la variabilidad cambian bruscamente, se observarán puntos extremos fuera de límites de control que se interpretan de acuerdo con la tabla siguiente:

	Gráfico $\overline{x}$	**Gráfico** R o s
Cambio en la media	Valor extremo	
Cambio en la dispers.	Valor extremo	Valor extremo

Un desplazamiento de la media del proceso producirá valores extremos en el gráfico de medias, pero no afectará a la dispersión del proceso, que continuará reflejando estado de control. Sin embargo, un cambio de la variabilidad puede generar puntos extremos tanto en el gráfico de la dispersión como en el de la media (ya que éstas tendrán entonces mayor variabilidad que la prevista en el gráfico) (véase la figura 13.10).

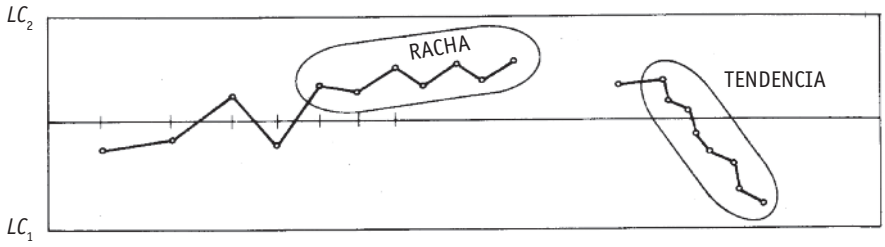
Figura 13.10 Efectos de cambios en la media y desviación



b) Tendencias en los puntos o rachas

Si el desplazamiento (de μ o σ) es paulatino a lo largo del tiempo (por desgaste de una herramienta, etc.), este cambio se detectará por un alineamiento de los puntos. En general 7 puntos consecutivos por encima o debajo de la media, o en orden creciente o decreciente, se consideran indicativos de anormalidad (figura 13.11), ya que la probabilidad de que esta configuración aparezca por azar es $(1/2)^7$, aproximadamente 3 entre mil.

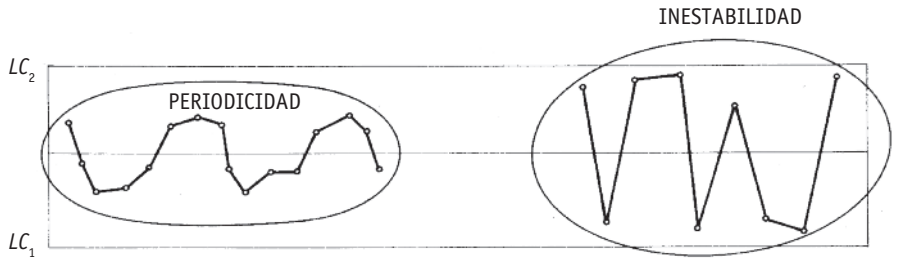
Figura 13.11 7 puntos con tendencia indican anormalidad



c) Periodicidades

Las diferencias entre turnos o en la calidad de la materia prima ocasionarán a veces gráficas con periodicidad en forma de ciclos, manifiesta en la sucesión de picos y valles (figura 13.12).

Figura 13.12 Periodicidades e inestabilidad



d) Inestabilidad

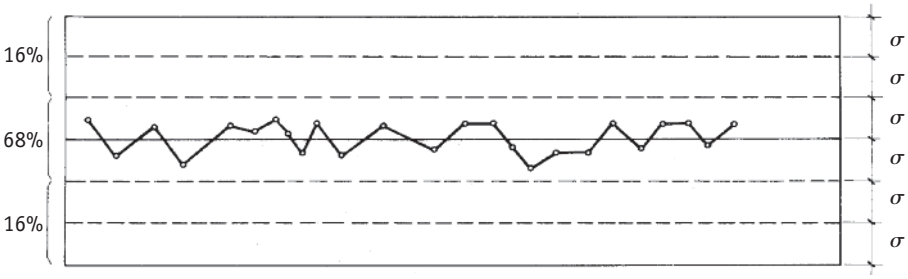
Se denomina inestabilidad a la presencia de grandes fluctuaciones que pueden producir uno o más puntos fuera de los límites de control. Este comportamiento puede ser debido a un sobreajuste de la máquina, a diferentes materiales mezclados en el almacén o a falta de entrenamiento del trabajador que controla el proceso.

e) Sobreestabilidad

Ocurre este fenómeno cuando la variabilidad de las muestras es menor que la esperada. Es importante identificar esta situación, ya que el análisis de las causas que la producen supone una oportunidad de reducir la variabilidad del proceso y aumentar su capacidad.

Para identificar este estado conviene situar en el gráfico dos líneas a cada lado de la línea central que dividan el intervalo de control en 6 partes iguales. En condiciones normales, el 68% de los puntos deberían estar entre las dos centrales y el 34% entre las siguientes.

Figura 13.13 Sobreestabilidad



Una acumulación de puntos en la zona central (figura 13.13) indica que los límites de control están mal calculados, que se han tomado incorrectamente los datos o que se ha producido un cambio positivo temporal en el proceso cuya causa debe investigarse.

13.6 Intervalos de tolerancia

Se define el *intervalo de tolerancia* para una caraterística de calidad, x, como el conjunto de valores de esta característica que se consideran admisibles. Tradicionalmente, si en una desviación del valor objetivo μ era mayor que L hacia el producto defectuoso, el intervalo de tolerancia se fijaba como $\mu \pm L$, y todos los elementos con medidas incluidas en este intervalo se consideraban igualmente buenos.

Este enfoque tiene dos inconvenientes. El primero es no considerar el coste de falta de calidad para el usuario que supone una desviación del valor nominal. Supongamos, por ejemplo, un circuito eléctrico que se instala en televisores y tostadores. Este circuito es defectuoso cuando su resistencia está fuera del intervalo $\mu \pm L$. Los circuitos defectuosos producen averías, cuya reparación cuesta 10.000 pesetas en el televisor, pero sólo 500 pesetas en el tostador. Aunque en ambos casos las resistencias son técnicamente defectuosas fuera de $\mu \pm L$, es razonable esperar que los límites de tolerancia deberán ser más estrechos para los circuitos que se instalen en televisores que para los que se instalan en tostadores.

En segundo lugar, el enfoque tradicional considera dos unidades cuyas características de calidad están incluidas en el intervalo de tolerancia como igualmente buenas. Sin embargo, cuanto mayor sea la desviación del valor nominal, peor será, en general, el funcionamiento de la unidad. Si el objetivo de la fabricación es conseguir el valor μ, cualquier desviación, aunque sea pequeña, supone siempre una pérdida de calidad que se traduce en un coste para el usuario. Taguchi, un ingeniero japonés, ha argumentado que los productos fabricados con intervalos de tolerancia que no tienen en cuenta las consecuencias de los errores de fabricación para el cliente no podrán sobrevivir en un mercado competitivo. Este autor sostiene que los intervalos de tolerancia deben establecerse teniendo en cuenta los costes para el cliente e igualando estos costes a los de la empresa que fabrica el producto.

13.6.1 La función de costes para el cliente

Para ilustrar cómo llevar a la práctica esta idea, consideremos el ejemplo simple de una pieza que debe encajar en otra y cuya característica de calidad es medir exactamente μ, valor que garantiza un acoplamiento óptimo.

Una pieza que mide x tendrá una falta de calidad $(x - \mu)$ que se traducirá en un coste para el usuario al no funcionar el acople en condiciones óptimas. Supongamos que, para el cliente o usuario:

a) Las desviaciones por exceso tienen el mismo efecto que por defecto.
b) Pequeñas desviaciones tienen un coste muy pequeño, pero éste aumenta rápidamente para desviaciones grandes.

Además, cuando la desviación del valor nominal sea suficientemente grande, por ejemplo, igual o mayor que M, lo que supone que x está fuera del intervalo $\mu \pm M$, el producto será inservible.

Estas hipótesis indican que el coste que para el usuario tiene una calidad x puede aproximadamente representarse mediante una función de coste del tipo:

$$C(x) = K(x - \mu)^2 \tag{13.3}$$

donde K es una constante que se determina fijando un punto de dicha curva. Una forma fácil de hacerlo es indicar el coste de reposición de un elemento defectuoso. Sea C_C el coste para el cliente de reponer una unidad con característica de calidad $x = \mu + M$ y que, por tanto, es defectuosa. Sustituyendo los valores $x = \mu + M$ y $C(\mu + M) = C_C$ en la ecuación (13.3), obtenemos que la constante debe verificar:

$$K = C_C / M^2$$

Por tanto, la función del coste para los clientes o usuarios de una pieza de dimensiones x es:

$$\boxed{C(x) = C_C \left(\frac{x - \mu}{M} \right)^2} \tag{13.4}$$

Por ejemplo, si la longitud de una pieza debe ser 625 mm (valor de μ), es defectuosa si se desvía más de 3 mm de este valor, ($M = 3$), y el coste de reposición para el usuario de una pieza defectuosa, C_C, es de 40 euros, la función de coste para nuestros clientes, o función de coste social, es:

$$C(x) = 40 \left(\frac{x - 625}{3} \right)^2$$

Por ejemplo, el coste para el usuario de una pieza aceptable, pero que mide 626 mm, un mm más que el valor nominal, es

$$C(626) = 40\left(\frac{1}{3}\right)^2 = 4{,}44 \text{ euros}$$

13.6.2 La determinación de tolerancias justas para el cliente

Diremos que las tolerancias son justas para el cliente cuando el fabricante ha fijado el intervalo de tolerancia igualando sus costes de reponer un elemento defectuoso a los costes que este elemento defectuoso produce al cliente. Para ello, si su coste de reponer una pieza defectuosa es *Cf*, el intervalo de tolerancia se calcula sustituyendo este coste del fabricante en la ecuación de costes para el cliente:

$$Cf = \frac{C_C}{M^2}(x - \mu)^2$$

y obteniendo la desviación, $(x - \mu)$, asociada a este coste, que será:

$$L = (x - \mu) = \pm M\sqrt{\frac{C_f}{C_c}} \qquad (13.5)$$

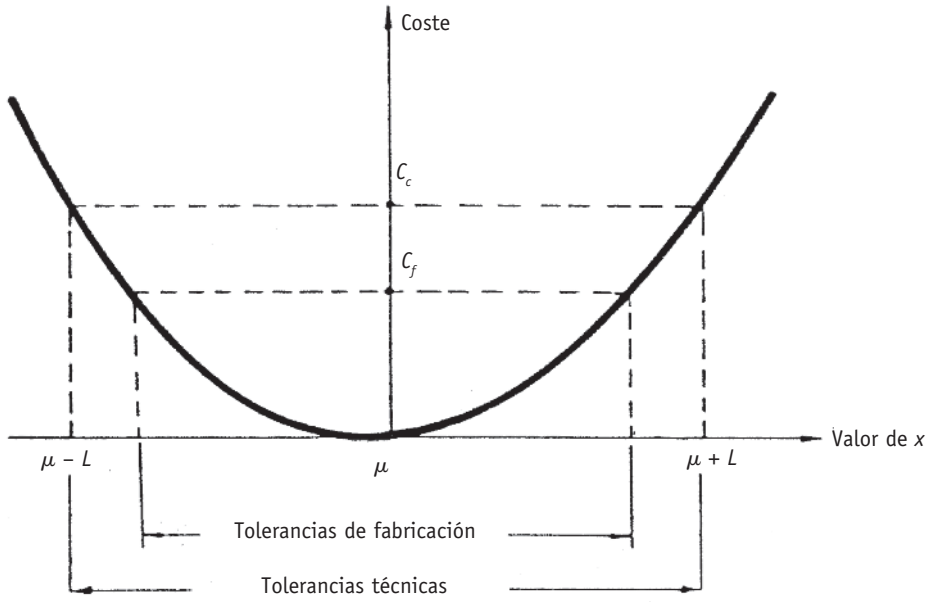
y éste debe ser el intervalo de tolerancia en la fabricación. En otros términos el fabricante debe determinar las tolerancias utilizando sus costes en la ecuación del cliente. De esta manera los costes de no calidad se reparten equitativamente entre ambos y se obtiene una solución de equilibrio. Por ejemplo, supongamos que en la función de costes del ejemplo anterior el coste de reposición de la pieza para el fabricante es de 5 euros. Entonces unas tolerancias de fabricación justa para el cliente son:

$$L = \pm 3\sqrt{\frac{5}{40}} = \pm 1{,}06$$

y, por tanto, serán inaceptables en la fabricación aquellas piezas con longitud fuera del intervalo 625 ± 1,06 mm. La figura 13.2 ilustra esta situación.

Cuando el coste de reposición para el usuario es poco mayor que para el fabricante, el intervalo de tolerancia resultante, $\mu \pm L$, será muy próximo a $\mu \pm M$; sin embargo, cuando el coste de reposición para el usuario es mucho mayor que para el fabricante —como ocurrirá siempre que el fallo de un elemento haga necesario sustituir un componente más amplio y costoso—, el intervalo de tolerancia para la fabricación será de amplitud mucho menor que 2*M*.

Figura 13.14 Fijación de las tolerancias para un producto



13.6.3 El coste de no calidad

El coste social esperado debido a la falta de calidad en la fabricación será el promedio de los costes para el cliente, es decir:

$$CT = E[C(x)] = \frac{C_c}{M^2} E(x - \mu)^2 = C_C \frac{\sigma^2}{M^2}$$

donde σ^2 es la varianza de la fabricación.

Por ejemplo, supongamos que en el ejemplo que estamos considerando la desviación típica de la fabricación es de 0,75 mm. El coste esperado de lanzar una pieza al mercado en este caso es

$$CT = 40 \left(\frac{0{,}75}{3} \right)^2 = 2{,}5 \text{ euros}$$

Supongamos que mejoramos el proceso de manera que la desviación típica de la fabricación se reduce a 0,5. El coste de no calidad será ahora

$$CT = 40 \left(\frac{0{,}5}{3} \right)^2 = 1{,}1 \text{ euros}$$

A igualdad de costes el cliente en un mercado abierto preferirá los productos con menor coste de no calidad. *Reducir la variabilidad de la fabricación es reducir los costes sociales por falta de calidad*.

13.7 El concepto de capacidad y su importancia

Cuando la característica de calidad es una medida continua, hemos visto que la falta de calidad depende de la variabilidad del proceso. Cuando el proceso está bajo control, es esperable que, según el teorema central del límite, la distribución de los valores de las características siga una distribución normal. En efecto, la variabilidad será debida a la suma de muchas causas independientes, cada una produciendo un efecto pequeño. Entonces la gran mayoría de las unidades fabricadas en condiciones de control (exactamente el 99,7%) se encuentran en un intervalo de amplitud 6σ (siendo σ la desviación típica de la distribución normal que define la fabricación). A este intervalo se le denomina intervalo de tolerancias naturales o intrínsecas del proceso. Por esta razón, cuando la característica de calidad es una medida, se define la capacidad del proceso como seis veces la desviación típica de esta característica en la producción, cuando el proceso está en condiciones de control estadístico. En consecuencia:

$$\text{Capacidad} = 6\sigma$$

El conocimiento de la capacidad de un proceso es imprescindible para juzgar su adecuación para la fabricación de productos con especificaciones y tolerancias dadas. Estas tolerancias, fijadas como hemos indicado en la sección anterior, reflejan la adecuación del producto al fin para el que está concebido.

13.7.1 Índice de capacidad

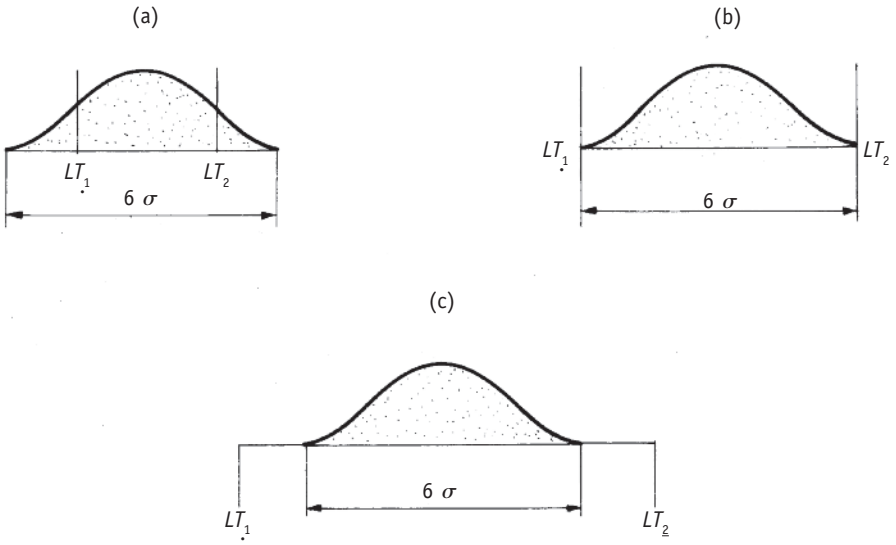
Sean (LT_1, LT_2) las tolerancias y supongamos que la media del proceso puede centrarse en $(LT_1 + LT_2)/2$. Entonces, el índice de capacidad del proceso C_p se define por:

$$C_p = \frac{(LT_2 - LT_1)}{6\sigma} \tag{13.6}$$

Suponiendo que la distribución de la característica medible es normal, según los valores de este índice podemos encontrarnos en alguno de los tres casos de la figura 13.15.

En el caso (a), $C_p < 1$ o $LT_2 - LT_1 < 6\sigma$, el proceso fabricará una proporción de defectuosos tanto más alta cuanto menor sea el índice de capacidad, no siendo capaz de cumplir las especificaciones fijadas. Habrá que actuar sobre el proceso tratando de disminuir la variabilidad no asignable, lo que requiere cambios en el proceso o en el producto. En caso contrario, la fabricación tendrá que someterse a un control muy frecuente y riguroso, para evitar que cualquier pequeño desajuste aumente más todavía el número de defectuosos. Esto supondrá un alto coste de muestreo y de reprocesar las unidades defectuosas. Nunca, en este caso, conviene modificar las tolerancias, que deben basarse en la adecuación al uso del producto y en las consideraciones económicas expuestas en la sección anterior y no en la capacidad del proceso.

Figura 13.15 Capacidad y tolerancias



En el caso (b), $C_p \approx 1$, el proceso fabricará aproximadamente un 0,3% de defectuosos. En el pasado, cuando un proceso cumplía esta condición, se le consideraba justamente apto para la fabricación, pero en la actualidad esta cantidad de defectos puede ser inaceptable en determinados productos en que los defectuosos se cuentan en tantos por millón. También en este caso el control requerido es muy estricto, ya que pequeños desplazamientos de la media aumentarán mucho la proporción de elementos defectuosos.

En el caso (c), el proceso fabricará una proporción de defectos muy pequeña. En general, diremos que el proceso es adecuado, aunque de nuevo el

concepto de pocos defectos es relativo al sector industrial. En este caso sólo es necesario supervisar el proceso para evitar desviaciones acusadas del estado de control.

Además de predecir el porcentaje de defectos, el estudio de la capacidad de un proceso es importante para:

1. Elegir entre procesos alternativos. La capacidad debe sopesarse con los costes de funcionamiento y de retroceso de defectos, para elegir el proceso productivo más adecuado entre los existentes.
2. Establecer un sistema de control de calidad durante la fabricación para detectar cuándo el proceso deja de estar en estado de control y tomar medidas correctoras.
3. Este control se realizará tomando cada cierto tiempo muestras de la fabricación y comprobando si éstas indican que el proceso se encuentra en estado de control. Aunque la frecuencia de muestreo depende, en última instancia, de criterios económicos, se pueden establecer las reglas de la tabla 13.4, de manera indicativa.
4. Indicar un punto de partida para la mejora del proceso: una medida objetiva de la eficacia de políticas de mejora es el aumento de la capacidad del proceso.

Tabla 13.4 Relación entre índice de capacidad y frecuencia de inspección (valores promedios aproximados que pueden variar mucho con la frecuencia de producción)

Índice de capacidad	Frecuencia de inspección
< 1	— Todas las unidades.
$1 < C_p < 1{,}4$	— Intensiva (cada 15 o 30 minutos).
$1{,}4 < C_p < 1{,}7$	— Moderada (cada hora).
$1{,}7 < C_p < 2$	— Cada 2 horas.
$2 < C_p$	— Depende de la frecuencia de causas anómalas.

Señalemos por último que la capacidad de un proceso se define en estado de control, por lo que un proceso con $C_p > 1$ puede producir en un momento dado un alto número de defectos si la materia prima es de calidad inferior, si la media no se ajusta adecuadamente en el centro de las tolerancias, si no recibe el mantenimiento adecuado y, en general, si no se mantiene en estado de control.

Ejemplo 13.3

La desviación típica de un proceso de fabricación de componentes electrónicos es 0,05 V. Se trata de fabricar dos componentes A y B que se consideran defectuosos si su tensión varía más de 0,5 V. Calcular el índice de capacidad del proceso si el coste de fabricación de ambos es el mismo, 50 euros, pero el coste de fallo para el usuario es en A de 80 euros y en B de 250 euros.

Suponiendo una función de coste cuadrática, el coste para el usuario será:

$$A: C(x) = \frac{80}{0{,}25}(x_A - \mu)^2; \qquad B: C(x) = \frac{250}{0{,}25}(x_B - \mu)^2;$$

por tanto, para fabricar el A, el intervalo de tolerancias debe ser, por (13.3):

$$x_A - \mu = \pm\, 0{,}5 \cdot (50/80)^{0{,}5} = 0{,}395$$

siendo el índice de capacidad:

$$IC_A = \frac{0{,}79}{0{,}30} = 2{,}64$$

y el proceso es muy capaz de cumplir con los requisitos. Para el segundo:

$$x_B - \mu = \pm\, 0{,}5 \cdot (50/250)^{0{,}5} = 0{,}224$$

con índice de capacidad:

$$IC_B = \frac{0{,}448}{0{,}30} = 1{,}49$$

y, por tanto, el proceso es adecuado y requiere una frecuencia de inspección moderada.

13.7.2 Un indicador alternativo de capacidad

A veces no es factible situar la media del proceso en el centro del intervalo, y es muy frecuente que los procesos tengan desplazamientos del valor medio de manera que la distribución de la característica de calidad no esté

en el centro del intervalo de tolerancia. En estos casos el índice C_p no es adecuado y en su lugar se utiliza el índice corregido C_{pk}. Este índice se define para tener en cuenta cambios en la media como sigue. Sea $m = (LT_2 + LT_1)/2$ el centro del intervalo de tolerancia o valor nominal. El índice se define como:

$$C_{pk} = mínimo \left\{ \frac{LT_2 - \mu}{3\sigma}, \frac{\mu - LT_1}{3\sigma} \right\}$$

donde μ es la media de la variable de calidad que puede no coincidir con el valor nominal. Supongamos que el valor nominal m es mayor que μ. Entonces, como μ está más cerca del límite inferior que del superior, $LT_2 - \mu > \mu - LT_1$, y tendremos que:

$$C_{pk} = \frac{\mu - LT_1}{3\sigma}$$

y restando y sumando m este índice puede escribirse

$$C_{pk} = \frac{m - LT_1 - (m - \mu)}{3\sigma} = C_p - \left(\frac{m - \mu}{3\sigma} \right)$$

Como estamos suponiendo $m \geq \mu$ es claro que $C_{pk} \leq C_p$. Por otro lado si $m = \mu$, entonces $C_{pk} = C_p$. Vamos a expresar la relación entre estos dos índices en función del desplazamiento de la media. Una forma de medir el desplazamiento $m - \mu$ (suponemos $m \geq \mu$) es compararlo con la distancia entre el valor nominal y el límite del intervalo de tolerancia. Para ello definimos la constante positiva

$$w = \frac{m - \mu}{(LT_2 - LT_1)/2}$$

entonces se verifica

$$\left(\frac{m - \mu}{3\sigma} \right) = wCp$$

y la relación entre los índices es

$$\boxed{C_{pk} = C_p(1 - w)}$$

Concluimos que cuando w es grande, entonces $C_{pk} << C_p$.

Ejemplo 13.4

Un proceso con índice de capacidad $C_p = 4/3 = 1{,}33$ tiene frecuentes desplazamientos de la media de hasta $1{,}5\sigma$ del valor nominal. Calcular el índice de capacidad corregido C_{pk}.

Si el desplazamiento de la media es $\mu = m + 1{,}5\sigma$, entonces, como $(LT_2 - LT_1) = C_p \cdot 6\sigma = 24\sigma/3 = 8\sigma$, tenemos que

$$w = \frac{1{,}5\sigma}{4\sigma} = 0{,}375$$

y entonces

$$\boxed{C_{pk} = 1{,}33(1 - 0{,}375) = 0{,}83}$$

Vemos que hay una disminución efectiva de la capacidad. Sin desplazamientos en la media el proceso produciría del orden de 62 piezas defectuosas por millón. En efecto, $P(z \leq 4) = 0{,}99996833$, y $2P(z > 4) = 2 \times 0{,}00003167$. Sin embargo, un deplazamiento de $1{,}5\sigma$ produce que la media esté a sólo $2{,}5\sigma$ del límite de tolerancia, y esto supone que la probabilidad de defectuoso pasará a ser $1 - 0{,}99379 = 0{,}00621$, es decir, de 6.210 por millón.

Una advertencia. Advertimos al lector que un error frecuente al establecer un sistema de control es hacer coincidir en el gráfico de control las líneas límites con las tolerancias del producto. Este error es especialmente grave cuando el índice de capacidad es menor que uno: entonces, aunque el proceso esté bajo control, obtendremos de vez en cuando valores fuera de los límites, lo que puede llevarnos a sobreajustar erróneamente el proceso y aumentar, en consecuencia, la proporción de defectuosos. El objetivo del control de fabricación es comprobar si el proceso permanece en estado de control, no si el producto está dentro de las tolerancias. La proporción de producto fuera de las tolerancias queda fijada, en estado de control, por el índice de capacidad del proceso.

13.8 El control de fabricación por atributos

13.8.1 Fundamentos

Ciertas características de calidad no están ligadas a ninguna variable numérica, sino a un atributo que un producto puede o no poseer: una lámpara se enciende o no, una tubería tiene o no fugas, un documento contiene o no errores, etc. En estos casos, se realiza un control por atributos.

Este tipo de control es también una alternativa al control por variables cuando, siendo ambos posibles, es necesario recoger datos rápidamente con poco coste: cualquier operario puede comprobar con un calibre pasa/no pasa si una pieza tiene una longitud dentro de tolerancias, pero se requiere cierta formación para calcular medias y desviaciones típicas de las longitudes de cinco piezas. En contrapartida, el control por atributos, al utilizar menos información, requiere tamaños muestrales sustancialmente más grandes que el control de variables.

El control por atributos se utiliza mucho en procesos administrativos y como primer paso al introducir métodos de control de calidad en una empresa: permite una identificación rápida de los problemas de calidad más urgentes y claves y sirve para definir aquellos parámetros de calidad del producto que presentan más problemas y donde conviene establecer un control de calidad por variables. Este control necesita únicamente un gráfico para controlar el proceso, en lugar de los dos que precisa el control por variables.

13.8.2 El estudio de capacidad

En el control por atributos se estudian muestras de tamaño n de elementos que se clasifican como aceptables o defectuosos. Llamaremos capacidad del proceso al valor $1 - p$, proporción de elementos aceptables fabricados en condiciones de control.

Si el proceso está en ese estado:

1) La proporción de elementos defectuosos fabricados, p, es estable a largo plazo, no tendiendo a aumentar ni a disminuir.
2) La producción de una pieza defectuosa en un momento dado es independiente de lo que haya ocurrido antes: las piezas defectuosas aparecen con la misma frecuencia después de piezas aceptables que después de defectuosas.

En estas condiciones, el número de piezas defectuosas en una muestra de tamaño n sigue la distribución binomial, con media np, y desviación típica $\sqrt{np(1-p)}$, quedando ambos parámetros determinados por la capacidad del proceso y el tamaño muestral.

Si en lugar de contar el número de defectuosos estudiamos la *fracción* defectuosa, p, ésta oscilará de una muestra a otra respecto al valor central p, con desviación típica $\sqrt{p(1-p)/n}$.

El proceso pasará a un estado de no control si la proporción de defectuosos varía (en general aumentará), lo que puede además ir unido a dependencia entre las apariciones de defectos: por ejemplo, tienden a darse en rachas; la aparición de uno hace más probable la aparición de otro, etc.

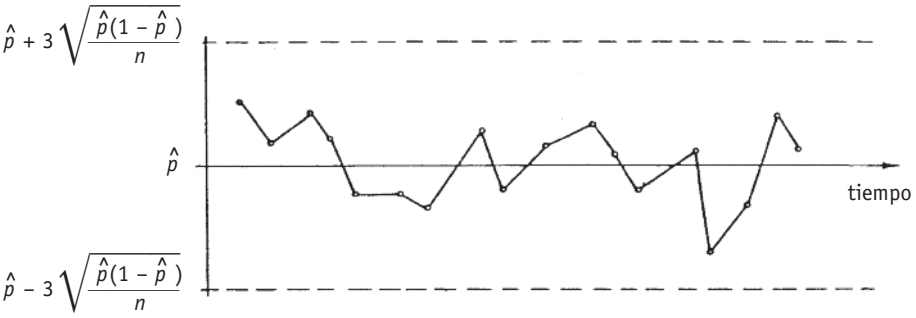
La capacidad de un proceso se estima por un procedimiento iterativo análogo al estudiado en el control por variables:

1) Tomar k muestras (k al menos 25) de n elementos (n mayor que 50) y contar el número de elementos defectuosos r en cada muestra. El período de recogida de datos debe ser suficiente para cubrir todas las causas de variabilidad que afectan al proceso en su funcionamiento habitual.
2) Estimar la proporción de elementos defectuosos por:

$$\hat{p} = \frac{r_1 + ... + r_k}{kn} = \frac{\text{número total de defectuosos}}{\text{número total de elementos}}$$

3) Comprobar que las k muestras son homogéneas respecto al valor de p, es decir, que el proceso ha permanecido en estado de control durante la recogida de información. Para ello, construiremos el gráfico:

Figura 13.16 Gráfico de proporción defectuosa



que contiene el tiempo en abscisas y en ordenadas la fracción defectuosa. Las líneas de control están situadas, como en el control por variables, a tres desviaciones típicas de la media.

Las etapas (2) y (3) se repiten eliminando aquellas muestras situadas fuera de los límites de control o que presentan tendencias, ciclos

o cualquier otra regularidad de las estudiadas en la sección de interpretación de gráficos de control, repitiendo el proceso hasta obtener una estimación basada en datos que, razonablemente, provienen del proceso en estado de control.

Ejemplo 13.5

Para estimar p en un control de fabricación por atributos se toman 20 muestras de 100 unidades, obteniendo los resultados de la tabla. Estimar la proporción defectuosa fabricada por el proceso y construir un gráfico de control para el número de defectos en muestras de 50 unidades. Estudiar la eficacia del gráfico calculando la probabilidad de detectar un punto fuera de los límites de control en función de la probabilidad de defecto.

Muestra	1	2	3	4	5	6	7	8	9	10
Defectos	0	0	3	2	0	4	1	1	2	0
Muestra	11	12	13	14	15	16	17	18	19	20
Defectos	8	1	2	3	1	0	3	2	1	1

El número total de defectos es 35; por tanto, dividiendo por el número de piezas observadas (2.000), se tiene:

$$\hat{p} = 0{,}0175$$

$$L_1 = 0{,}0175 + \frac{3}{\sqrt{n}}\sqrt{pq} = 0{,}056$$

$$L_2 = 0$$

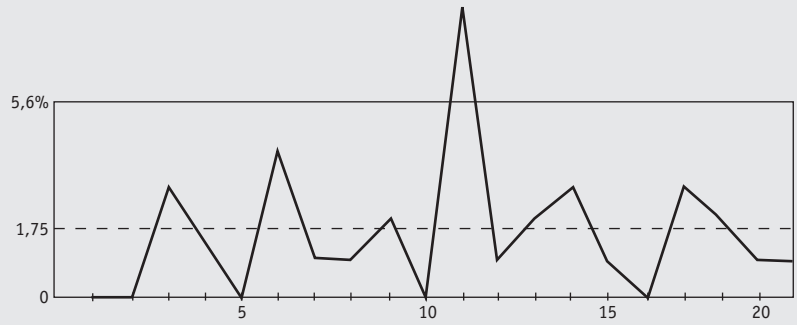
resultando la figura 13.17, donde los resultados se expresan en %.

El punto con 8 está fuera de control y lo eliminaremos. Con 19 muestras:

$$\hat{p} = \frac{27}{1.900} = 0{,}0142 \qquad L_1 = 0{,}0142 + \frac{3}{\sqrt{100}}\sqrt{pq} = 0{,}0497$$

y ahora todos los puntos están bajo control. El gráfico de control para $n = 50$ tendrá una línea central en $np = 50 \cdot 0{,}0142 = 0{,}71$ y una línea superior en $0{,}71 + 3\sqrt{0{,}71 \cdot 0{,}9868} = 3{,}22$. Por tanto, más de tres defectuosas en una muestra de tamaño 50 se considerará como anormal.

Figura 13.17 Gráfico de control para el ejemplo 13.4



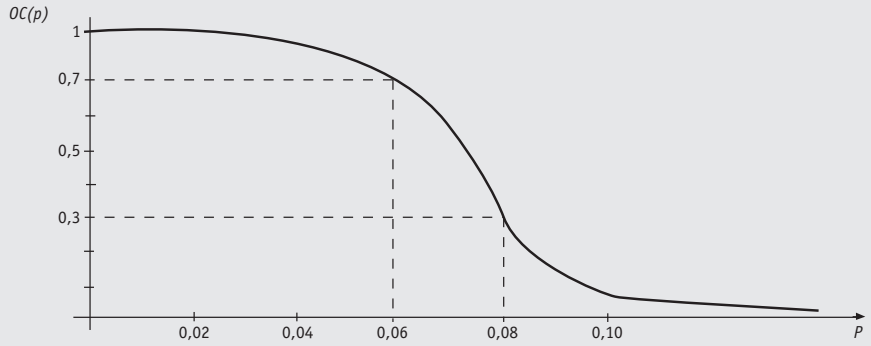
Para calcular la eficacia de este plan de control vamos a construir la curva que da la probabilidad de aceptar que el proceso está bajo control para cada valor de p (curva *OC,* sección 13.2). Sea x el número de defectos en la muestra. Entonces, definiendo

$$OC(p) = p(\text{estando de control}) = P(x > 4|p) = P(x \leq 3{,}5|p) =$$

$$= P\left(\frac{x - 50p}{\sqrt{50pq}} \leq \frac{3{,}5 - 50p}{\sqrt{50pq}}\right) = P[z \leq (3{,}5 - 50p)/\sqrt{50pq}].$$

y dando valores a p se obtiene una curva como la figura 13.18.

Figura 13.18



13.8.3 Gráficos de control

Los gráficos que se utilizan son el gráfico p, proporción observada de defectos, o el gráfico np, número de elementos defectuosos. El primero es preferido cuando n varía de una muestra a otra, mientras que el segundo se

utiliza con tamaño muestral fijo. Su interpretación es muy similar. Cuando un punto muestral salga fuera de los límites de control, las opciones posibles son:

a) El proceso ha variado, aumentando o disminuyendo (según el sentido del valor extremo) el valor de p.
b) El sistema de medición ha cambiado (el inspector o los criterios de medida).
c) Se ha cometido un error al calcular el valor de $\hat{p}$ en dicha muestra.
d) El proceso no ha variado, pero los límites de control son erróneos.
e) Nada ha cambiado, simplemente un suceso poco frecuente ha ocurrido.

La primera comprobación es descartar (b) y (c). Una vez comprobado este extremo, la información previa sobre el proceso determinará la elección entre las causas restantes: si se dispone de abundante información previa sobre la capacidad del proceso y podemos desechar (d) nos encontraremos con una de las situaciones siguientes:

a) La probabilidad de (e) es sólo de tres casos entre mil, por lo que normalmente será mucho más probable que el proceso se haya desajustado e investigaremos la causa.
b) El proceso se desajusta muy raramente, con lo que ambas causas son verosímiles. Entonces tomaremos nuevas muestras para ver si la falta de control se confirma.

13.9 El control de fabricación por números de defectos

13.9.1 Fundamentos

El control por atributos no resulta adecuado cuando los defectos no van asociados a unidades, sino que aparecen en un flujo continuo de producto (defectos en una plancha fotográfica, burbujas en un cristal, etc.). En estas situaciones, aunque es posible dividir el producto continuo fabricado en «elementos» de observación, resulta más conveniente y práctico considerar el carácter continuo de los elementos considerados. Por ejemplo, en la fabricación de rollos de papel podemos suponer que cada elemento son 10 cm de papel y contar cuántos elementos (porciones de 10 cm) tienen defecto, lo que permitiría utilizar un gráfico p o np. Sin embargo, es más conveniente considerar directamente el número de defectos por metro, rollo, etc.

El control por número de defectos debe también aplicarse cuando los elementos de la fabricación puedan tener varios defectos independientes y

la calidad dependa del número de éstos. Un análisis que clasifique estos elementos únicamente como defectuosos o no desperdiciaría esta información y sería inadecuado.

13.9.2 Estudios de capacidad y gráficos de control

Diremos que un proceso está bajo control cuando:

1) El proceso es estable y fabrica un número medio de defectos por unidad de longitud, área, etc., constante, que llamaremos m.
2) Los defectos aparecen independientemente unos de otros.

En estas condiciones, el número de defectos por unidad de observación sigue una distribución de Poisson. La capacidad del proceso se define por m, número medio de defectos. Por tanto, si tomamos muestras del mismo tamaño n (longitud, área, etc.), el número medio de defectos oscilará alrededor del valor medio, m, para dicho tamaño, con desviación típica $\sqrt{m/n}$. Para estimar m, se realiza el proceso iterativo expuesto en casos anteriores:

1) Tomar k muestras de tamaño (longitud, área, etc.) n y determinar el número de defectos en cada muestra, $c_1 \ldots c_k$.
2) Estimar m mediante

$$\hat{m} = \frac{c_1 + \ldots + c_k}{nk}$$

3) Construir alguno de los gráficos de control que describimos a continuación partiendo de que $m = \hat{m}$ y utilizarlo para eliminar datos fuera de control, repitiendo después los pasos anteriores.

Existen dos tipos de gráficos de control que corresponden a los casos siguientes:

a) La muestra observada es siempre del mismo tamaño n: si el proceso es continuo, contiene siempre la misma longitud, área, volumen; si es discreto, el mismo número de elementos. El gráfico adecuado es entonces el del número de defectos observados en cada muestra, con línea central $c = nm$ y líneas de control $c \pm 3\sqrt{c}$.
b) Si las muestras no son del mismo tamaño, sean $n_1, \ldots, n_k$ los tamaños muestrales. Entonces m se estima por $\Sigma c_i / \Sigma n_i$ y los límites de control no son fijos, sino que dependen del tamaño de muestra. La línea central será $\hat{m}$ y los límites de control para la muestra n_i serán $\hat{m} \pm 3\sqrt{\hat{m}/n_i}$.

Ejemplo 13.6

Se han tomado muestras de cable y analizado el número de defectos. La longitud de la muestra no es siempre idéntica, como indica la tabla.

N.º muestra	1	2	3	4	5	6	7	8	9	10
Longitud (m)	1	1	1	1	1,5	1,5	1,5	1,5	1,5	1,5
N.º defectos	3	4	3	3	5	6	7	4	3	4
N.º muestra	11	12	13	14	15	16	17	18	19	20
Longitud (m)	1,5	1,5	1,5	1,5	1,5	1,8	1,8	1,8	1,8	1,8
N.º defectos	5	2	4	3	6	4	7	8	4	8

Estudiar si el proceso está bajo control.
La media de defectos por *m* de longitud observada es:

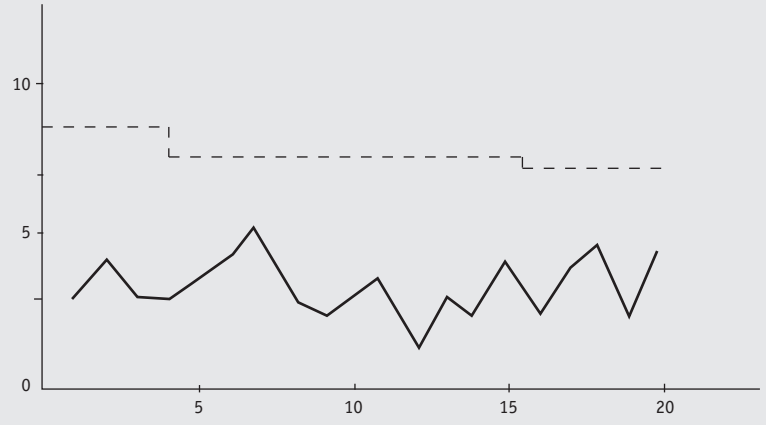
$$\hat{m} = \frac{3 + 4 + 3 + 3 + 5 + \ldots + 6 + 4 + \ldots + 8}{1 \cdot 4 + 1{,}5 \cdot 11 + 1{,}8 \cdot 5} = \frac{93}{29{,}5} = 3{,}15$$

y los límites de control:

a) Para tamaño 1 : $3{,}15 \pm 3\sqrt{3{,}15} = 3{,}15 \pm 5{,}32$; (0; 8,47).
b) Para tamaño 1,5 : $3{,}15 \pm 3\sqrt{3{,}15/1{,}5} = 3{,}15 \pm 4{,}35$; (0; 7,5).
c) Para tamaño 1,8 : $3{,}15 \pm 3\sqrt{3{,}15/1{,}8} = 3{,}15 \pm 3{,}97$; (0; 7,12).

La figura 13.19 presenta el gráfico de control.

Figura 13.19 Gráfico de control para *m*



13.10 Los gráficos de control como herramientas de mejora del proceso

13.10.1 La mejora de procesos

El control de procesos mediante gráficos de control pretende:

a) Asegurar que el proceso está en estado de control.
b) Aprender sobre el proceso identificando causas que influyen sobre la media y la variabilidad, midiendo sus efectos y aprovechándolos. En este sentido se convierten en una herramienta para mejorar el proceso.

El enfoque tradicional del control de calidad se ha centrado en el primer aspecto, pero progresivamente, a la vista de la experiencia de Japón y de las enseñanzas de Box, Deming y Jurán, entre otros, las funciones de aprendizaje y mejora del proceso han ido adquiriendo el papel central.

Esta evolución corresponde a distintos grados de madurez en el estudio del proceso. Asegurar el estado de control es el primer paso imprescindible para cualquier estudio de mejora, y fue la motivación principal de Shewhart cuando en los años veinte introdujo los gráficos de control. Como un proceso en estado de control estadístico tiene un funcionamiento predecible, esta situación permite una adecuada planificación de la producción y establece una base para medir con rapidez, precisión y objetividad los cambios en el sistema.

El segundo paso es investigar las causas que producen valores fuera de control: cuando la aparición de un punto atípico en un gráfico de control supone buscar una causa, corregir sus defectos mediante un ajuste del proceso y tomar medida para evitar su aparición futura, estamos utilizando los gráficos de control para aprender sobre el proceso y mejorarlo. Especial relevancia debe darse a las desviaciones *positivas*, que indican el efecto de causas que *mejoran* el proceso (cambios de materia prima, de política de mantenimiento preventivo, etc.). Su análisis e identificación es tanto o más importante que el de las causas negativas.

De acuerdo con Jurán (1983), el 85% de los problemas de calidad en las empresas dependen del proceso, y sólo un 15% de causas asignables. Deming asegura que el 94% depende del proceso y son, por tanto, responsabilidad de la dirección, mientras que sólo un 6% corresponden a causas especiales. Por tanto, el objetivo de mejora continua del proceso, modificándolo activamente y midiendo sus resultados con los gráficos de control, resulta especialmente clave. En particular, la experiencia muestra que los siguientes aspectos son causa de alta variabilidad —baja capacidad— en muchos procesos:

— Mal diseño del producto.
— Falta de atención a la formación de los trabajadores en métodos estadísticos simples.
— Mala supervisión, con estándares que inducen a errores y provocan conflictos frecuentes entre trabajadores e inspectores.
— Materia prima de calidad deficiente.
— Falta de motivación del trabajador para realizar un buen trabajo.
— Mezcla de materias primas de características de calidad distintas.

Un tercer paso en la mejora de los procesos es, en lugar de observarlos pasivamente, experimentar sobre ellos para descubrir posibilidades de mejora. Una forma eficaz de realizar esta experimentación es utilizar la teoría estadística de diseño de experimentos, que estudiaremos en el segundo volumen de este libro y que permite medir conjuntamente los efectos de muchas variables sobre la característica de calidad del proceso. La utilización de esta idea ha llevado al enfoque seis sigma, que comentamos a continuación.

13.10.2 El enfoque seis sigma

Si conseguimos mejorar de forma sistemática un proceso eliminando sucesivamente las fuentes de variación, podemos aumentar la capacidad. El enfoque seis sigma preconiza conseguir un índice de capacidad de 2, de manera que los intervalos de tolerancia estén a seis desviaciones típicas (de ahí el nombre de seis sigma) del valor nominal. De esta manera, si se producen cambios en la media del proceso menores de 1,5 desviaciones típicas, que son típicamente difíciles de detectar rápidamente con un gráfico de control, nos aseguramos de que la mayoría de la fabricación estará dentro de los límites de tolerancia.

La experiencia acumulada de muchas empresas muestra que son muy frecuentes desplazamientos pequeños de la media que sería muy difícil y costoso detectar. El enfoque seis sigma parte de esta idea, y en lugar de utilizar como índice de capacidad C_p, que supone que la media del proceso está fija en el valor nominal, piensa en términos del índice C_{pk}, que tiene en cuenta los desplazamientos de la media. Se trata de mantener $C_{pk} \geq 1{,}5$ para asegurar que el número de defectos se mantenga en valores del orden de 3 por millón.

Ejercicios 13.1

13.1.1. Para determinar la capacidad de un proceso se han tomado 15 muestras de tamaño 5, con los resultados siguientes:

N.º muestra		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
D	1	188	001	130	184	099	124	025	113	080	231	126	177	113	157	000
A	2	038	067	195	177	108	043	053	438	113	017	039	009	181	012	279
T	3	064	034	000	178	351	028	145	278	017	095	050	062	009	043	100
O	4	186	113	243	278	012	070	187	235	082	169	291	002	010	198	123
S	5	002	023	206	060	164	188	034	310	001	132	337	054	061	062	121
	$\bar{x}$	95,6	47,6	154,8	175,4	146,8	90,6	88,8	274,8	58,6	128,8	168,6	60,8	74,8	94,4	124,6
	R	186	112	243	218	339	160	162	335	112	214	334	175	171	186	279

$\bar{\bar{x}} = 119{,}00$
$\bar{R} = 215{,}07$

Se pide:

a) Construir gráficos para la media y el rango.
b) Estimar la capacidad del proceso.
c) Comprobar la normalidad de los datos.

13.1.2. En una fabricación de resistores se toman 20 muestras de tamaño 4, obteniendo $\Sigma \bar{x}_i = 8620$, $\Sigma R_i = 910$, siendo $\bar{x}_i$ y R_i la media y rango de cada muestra. Se pide:

a) Construir gráficos de control para la media y el rango.
b) Estimar σ.
c) Calcular el índice de capacidad si los límites de tolerancia son 430 ± 30 y la proporción de elementos defectuosos en la fabricación.

13.1.3. El control de un proceso por variables se realiza tomando cada hora una muestra de 4 unidades y calculando $\bar{x}$ y s. Si se produce un desplazamiento en la media del proceso de $0{,}5\sigma$, sin modificarse la desviación típica σ, calcular el tiempo que se tardará en apreciar el cambio.

13.1.4. Un control de fabricación por variables se realiza tomando muestras de $n = 9$. Calcular la probabilidad de que una muestra al azar no salga fuera de los límites de control para la media —supuesto σ conocida— cuando la media del proceso se desplaza $k\sigma$, siendo σ constante, y representar estas probabilidades en función de k.

13.1.5. Se han tomado 20 muestras en un control de 200 elementos por atributos con los resultados siguientes:

N.º de muestra	1	2	3	4	5	6	7	8	9	1	0
N.º de defectos	20	15	16	18	23	22	20	18	15	4	0
N.º de muestra	11	12	13	14	15	16	17	18	19	2	0
N.º de defectos	13	33	46	38	29	26	27	17	12	1	4

Construir un gráfico de control y estimar la proporción defectuosa.

13.1.6. Se han observado 20 muestras de 3 m^2 de material textil con el siguiente resultado en defectos por metro cuadrado observado:

N.º de muestra	1	2	3	4	5	6	7	8	9	1	0
N.º de efectos	3	4	3	2	1	6	4	5	7	2	
N.º de muestra	11	12	13	14	15	16	17	18	19	2	0
N.º de defectos	2	6	3	4	7	3	2	4	1	3	

Estudiar si el proceso ha estado bajo control.

13.1.7. Si en el proceso anterior las cinco primeras muestras se hubiesen obtenido analizando 1 m^2 y las 15 siguientes 3 m^2, ¿qué diferencias aparecerían en el análisis?

13.1.8. Para estudiar el funcionamiento de un muelle de descarga se establece un gráfico de control. Los camiones llegan a una media de 6,2 a la hora.

a) Diseñe un gráfico de control para estudiar el sistema.
b) ¿Cuáles serían los límites de control para llegadas diarias? (8 horas de trabajo).

13.1.9. Para controlar de forma aproximada un generador de números aleatorios se dibuja la suma cada 100 dígitos consecutivos generados. Determinar los gráficos de control correspondientes.

13.11 El control de recepción

13.11.1 Planteamiento del problema

El control de recepción se aplica al recibir materias primas, materiales o productos intermedios que serán introducidos en la fabricación, para comprobar que cumplen las especificaciones de calidad. Este control puede realizarse sobre características medibles (control de recepción por variables), atributos o número de defectos. Vamos a ocuparnos únicamente del control de recepción por atributos, que es el más utilizado.

Al recibir un lote de unidades, el comprador puede optar por aceptarlo, inspeccionarlo al 100% o inspeccionarlo mediante muestreo. Sea N el tamaño del lote, $C(i)$ el coste de inspeccionar una unidad y $C(d)$ el coste de introducir en el proceso una unidad defectuosa. Entonces, llamando $p(d)$ a la proporción de defectuosos en el lote, tendremos:

$$\text{Coste de no inspeccionar} = C(d)\, p(d)\, N$$

$$\text{Coste de inspección al 100\%} = C(i)\, N$$

Por tanto, si $p(d)$ fuese conocida, la decisión sería:

$$\text{Inspección al 100\% si: } C(i)\, N < C(d)\, p(d)\, N$$

es decir:

$$p(d) > C(i)/C(d)$$

y no inspeccionar en otro caso. En la práctica, el valor de $p(d)$ es desconocido y se estima mediante muestreo. Si la proporción de defectuosos en el lote se estima mayor que $C(i)/C(d)$, el lote se rechaza o se inspecciona al 100%, sustituyendo todos los elementos defectuosos; en caso contrario el lote se acepta.

El control por muestreo es la única alternativa para estimar la calidad del producto recibido cuando:

1) Los ensayos son destructivos, como ocurre cuando la calidad se define por la resistencia a la rotura, la duración de vida, etc.
2) La inspección de cada unidad tiene un coste muy elevado o requiere un tiempo muy largo.
3) En grandes partidas de material en las que resulta inviable la inspección de todas ellas por su magnitud (tornillos, tuercas, etc.) o su longitud (papel, película fotográfica, textiles, etc.).

13.11.2 El control simple por atributos

Supongamos que se recibe un lote grande de productos y que el comprador considera que si la proporción de elementos defectuosos es igual o menor a un valor p_A, que llamaremos *nivel de calidad aceptable* (NCA y en inglés AQL, de *Aceptable Quality Level)*, el lote debe aceptarse. En caso contrario, el lote debería rechazarse. El valor p_A depende de criterios económicos y técnicos (cuando rechazar un lote supone inspeccionarlo al 100%, p_A puede ser igual a $C[i]/C[d]$). Se trata de diseñar un contraste de hipótesis para aceptar o rechazar el lote.

Diseñar un contraste requiere elegir un tamaño muestral, n, y un número de defectuosos c tal que:

> Si $x > c$ rechazaremos el lote
> Si $x \leq c$ aceptaremos

Llamando p a la proporción verdadera desconocida de defectos en el lote, al realizar el contraste podemos cometer dos tipos de errores: aceptar un lote que deberíamos rechazar ($p > p_A$) o rechazar un lote que deberíamos aceptar ($p \leq p_A$). Se define:

α = riesgo del vendedor = probabilidad de rechazar un lote con $p = p_A$

El valor α corresponde a la probabilidad de un error tipo I estudiada en el capítulo 5, ya que la hipótesis nula del contraste es que el lote es aceptable ($p = p_A$). Para encontrar un plan de muestreo, es decir, unos valores de n y c, consistentes con un valor dado de α, se utiliza la distribución binomial (suponemos que el tamaño del lote, N, es muy grande comparado con n, tamaño muestral, y por tanto la binomial es adecuada). Por ejemplo, fijando n y buscando el menor valor c de manera que:

$$\sum_{i=c+1}^{n} \binom{n}{i} p_A^i (1 - p_A)^{n-i} \leq \alpha$$

Definir adecuadamente el plan exige tener también en cuenta el error tipo II: aceptar un lote que debería rechazarse. Para controlar este error, se especifica un cierto valor p_R llamado *nivel de calidad rechazable* (NCR, en

inglés RQL o LTPD), tal que la probabilidad de aceptar lotes con calidad igual o menor que p_R debe de ser muy baja. Se define:

$$\beta = \text{riesgo del comprador} =$$

$$= \text{probabilidad de aceptar un lote con } p = p_R$$

Normalmente p_R/p_A está entre 4 y 10.

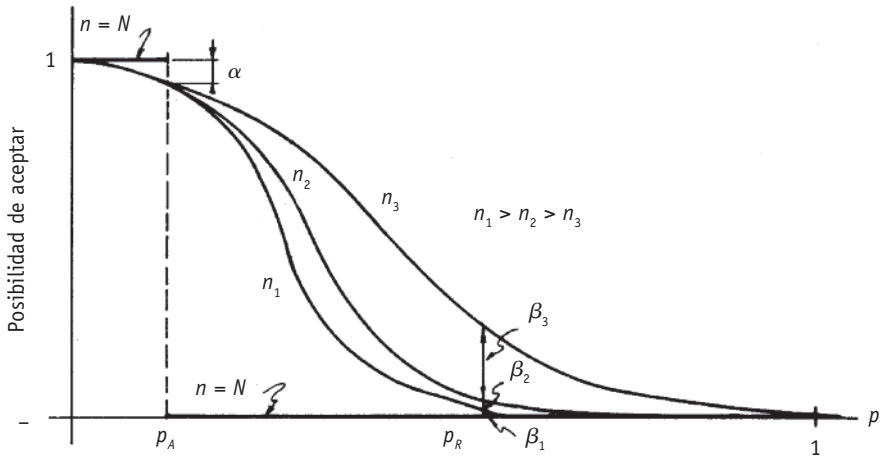
El plan de muestreo se determina fijando los cuatro valores (α, p_A), (β, p_R). En el control de recepción, para caracterizar el contraste se utiliza la *curva característica* del contraste, definida por:

$$\text{Curva característica} = OC(p) = 1 - \text{curva de potencia} = 1 - \text{Pot}(p)$$

La curva característica proporciona, por tanto, para cada valor de p, la probabilidad de aceptar un lote. Especificar (α, p_A) y (β, p_R) equivale a fijar dos puntos de la curva característica.

La figura 13.20 presenta distintas curvas características de planes de muestreo al aumentar el tamaño muestral. Cuando $n = N$ e inspeccionamos el lote al 100% tenemos la curva ideal, sin posibilidades de error. A medida que disminuye n, manteniendo α constante, aumenta el error β. Al fijar (α, p_A) y (β, p_R) el plan de muestreo queda determinado.

Figura 13.20 Curvas características



Ejemplo 13.7

Diseñar un plan de muestreo para lotes de 10.000 unidades, con los siguientes parámetros:

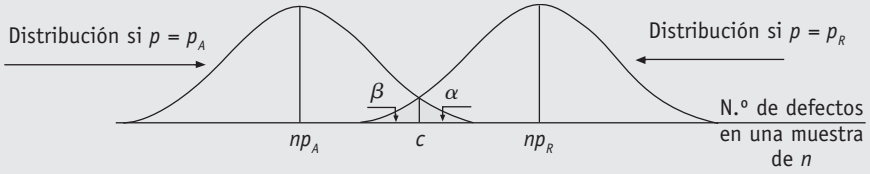
$$P_A = 0{,}02;\ \alpha = 0{,}05;\ p_R = 0{,}04;\ \beta = 0{,}05$$

La variable x, número de defectuosos, es aproximadamente Poisson con $\lambda = np$. Suponiendo n grande y aproximando por la normal, x será $N(np, \sqrt{np})$. Entonces, llamando c al valor de rechazo y p a la proporción defectuosa en el lote:

$$P(\text{rechazo}/p_A) = P(r > c|p = 0{,}02) = 0{,}05 = \alpha$$
$$P(\text{aceptar}/p_R) = P(r \leq c|p = 0{,}04) = 0{,}05 = \beta$$

La figura 13.21 ilustra la situación.

Figura 13.21 Cálculo de c conocidos (α, p_A) y (β, p_R)



Entonces, si z es una variable $N(0, 1)$:

$$c = np_A + z_\alpha \sqrt{np_A},\ \text{donde } z_\alpha \text{ se define por } P(z > z_\alpha) = \alpha$$
$$c = np_R + z_\beta \sqrt{np_R},\ \text{donde } z_\beta \text{ se define por } P(z < z_\beta) = \beta$$

Por tanto, igualando ambas expresiones de c y despejando n:

$$n = \left[\frac{z_\alpha \sqrt{p_A} - z_\beta \sqrt{p_R}}{p_R - p_A} \right]^2$$

En el ejemplo $z_\alpha = -z_\beta = 1{,}64$; sustituyendo:

$$n = [(1{,}64\sqrt{0{,}02} + 1{,}64\sqrt{0{,}04})/(0{,}04 - 0{,}02)]^2 = 783{,}8$$

Tomando $n = 784$ y sustituyendo en la expresión de c, se obtiene:

$$c = 784 \cdot 0{,}02 + 1{,}64\sqrt{0{,}02 \cdot 784} = 22{,}17$$

Por tanto, el plan que reúne estos requisitos es:

Si $c \geq 23$ rechazar.
Si $c \leq 22$ aceptar el lote.

13.11.3 Planes de muestreo

La determinación de un plan de muestreo a partir de dos puntos de la curva característica es laborioso; para simplificar esta tarea se han construido tablas que los proporcionan. Estos planes pueden clasificarse en:

a) Planes de aceptación/rechazo: los más conocidos son las normas japonesas (JIS Z 9002) y las norteamericanas (Military Standard), que se han convertido en normas internacionales (ISO) y españolas (UNE).
b) Planes de control rectificativo: se diferencian de los anteriores en que los lotes rechazados se inspeccionan al 100% sustituyendo los elementos defectuosos. Los más utilizados son debidos a Dodge-Romig.

Vamos a analizar estos planes.

13.11.4 Plan japonés JIS Z 9002

La tabla 13.5 reproduce un ejemplo de las tablas construidas para aplicarlo en el caso $\alpha = 0{,}05$, $\beta = 0{,}10$. Se fijan los valores P_A y P_R y la tabla proporciona los valores de n y c (en negritas). Por ejemplo, si $p_A = 2\%$ y $P_R = 10\%$, el plan es tomar 60 piezas y rechazar si el número de defectuosos es *mayor* que 3. Estos planes tienen, aproximadamente, una curva característica que pasa por los puntos fijados.

13.11.5 Plan Military-Standard (MIL-STD-105D; ISO 2859; UNE 66020)

Estos planes fueron desarrollados inicialmente por el ejército estadounidense durante la Segunda Guerra Mundial y han sufrido mejoras poste-

Tabla 13.5 Inspección por atributos, norma JIS Z 9002 (α = 0,05; β = 0,10)

p_R(%)	0,71 ~	0,91 ~	1,13 ~	1,41 ~	1,81 ~	2,25 ~	2,81 ~	3,56 ~
p_A(%)	0,90	1,12	1,40	1,80	2,24	2,80	3,55	4,50
0,090~0,112	*	400 **1**	↓	←	↓	→	60 **0**	50 **0**
0,113~0,140	*	↓	300 **1**	↓	←	↓	→	↓
0,141~0,180	*	500 **2**	↓	250 **1**	↓	←	↓	→
0,181~0,224	*	*	400 **2**	↓	200 **1**	↓	←	↓
0,225~0,280	*	*	500 **3**	300 **2**	↓	150 **1**	↓	←
0,281~0,355	*	*	*	400 **3**	250 **2**	↓	120 **1**	↓
0,356~0,450	*	*	*	500 **4**	300 **3**	200 **2**	↓	100 **1**
0,451~0,560	*	*	*	*	400 **4**	250 **3**	150 **2**	↓
0,561~0,710	*	*	*	*	500 **6**	300 **4**	200 **3**	120 **2**
0,711~0,900	*	*	*	*	*	400 **6**	250 **4**	150 **3**
0,901~1,12		*	*	*	*	*	300 **6**	200 **4**
1,13 ~1,40			*	*	*	*	500 **10**	250 **6**
1,41 ~1,80				*	*	*	*	400 **10**
1,81 ~2,24					*	*	*	*
2,25 ~2,80						*	*	*
2,81 ~3,55							*	*
3,56 ~4,50								*
4,51 ~5,60								
5,61 ~7,10								
7,11 ~9,00								
9,01 ~11,2								
p_A(%)	0,71 ~	0,91 ~	1,13 ~	1,41 ~	1,81 ~	2,25 ~	2,81 ~	3,56 ~
p_R(%)	0,90	1,12	1,40	1,80	2,24	2,80	3,55	4,50

(Use la primera columna en la dirección de la flecha, no hay métodos que satisfagan los requisitos en las columnas en blanco).
(Las tablas dan *n*, tipo de letra normal, y *c*, número máximo de defectuosos admisibles, en negrita).

4,51 ~ 5,60	5,61 ~ 7,10	7,11 ~ 9,00	9,01 ~ 11,2	11,3 ~ 14,0	14,1 ~ 18,0	18,1 ~ 22,4	22,5 ~ 28,0	28,1 ~ 35,5	P_R(%) / P_A(%)
←	↓	↓	←	↓	↓	↓	↓	↓	0,090~0,112
40 **0**	←	↓	↓	←	↓	↓	↓	↓	0,113~0,140
↑	30 **0**	←	↓	↓	←	↓	↓	↓	0,141~0,180
→	↑	25 **0**	←	↓	↓	←	↓	↓	0,181~0,224
↓	→	↑	20 **0**	←	↓	↓	←	↓	0,225~0,280
←	↓	→	↑	15 **0**	←	↓	↓	←	0,281~0,355
↓	←	↓	→	↑	15 **0**	←	↓	↓	0,356~0,450
80 **1**	↓	←	↓	→	↑	10 **0**	←	↓	0,451~0,560
↓	60 **1**	↓	←	↓	→	↑	7 **0**	←	0,561~0,710
100 **2**	↓	50 **1**	↓	←	↓	→	↑	5 **0**	0,711~0,900
120 **3**	80 **2**	↓	40 **1**	↓	←	↓	↑	↑	0,901~1,12
150 **4**	100 **3**	60 **2**	↓	30 **1**	↓	←	↓	↑	1,13 ~1,40
200 **6**	120 **4**	80 **3**	50 **2**	↓	25 **1**	↓	←	↓	1,41 ~1,80
300 **10**	150 **6**	100 **4**	60 **3**	40 **2**	↓	20 **1**	↓	←	1,81 ~2,24
*	250 **10**	120 **6**	70 **4**	50 **3**	30 **2**	↓	15 **1**	↓	2,25 ~2,80
*	*	200 **10**	100 **6**	60 **4**	40 **3**	25 **2**	↓	10 **1**	2,81 ~3,55
*	*	*	150 **10**	80 **6**	50 **4**	30 **3**	20 **2**	↓	3,56 ~4,50
*	*	*	*	120 **10**	60 **6**	40 **4**	25 **3**	15 **2**	4,51 ~5,60
	*	*	*	*	100 **10**	50 **6**	30 **4**	20 **3**	5,61 ~7,10
		*	*	*	*	70 **10**	40 **6**	25 **4**	7,11 ~9,00
			*	*	*	*	60 **10**	30 **6**	9,01 ~11,2
4,51 ~ 5,60	5,61 ~ 7,10	7,11 ~ 9,00	9,01 ~ 11,2	11,3 ~ 14,0	14,1 ~ 18,0	18,1 ~ 22,4	22,5 ~ 28,0	28,1 ~ 35,5	p_A(%) / p_R(%)

Tabla 13.6 Códigos para Military Standard

Tamaño del lote	Niveles especiales de inspección				Niveles normales de inspección		
	S-1	S-2	S-3	S-4	I	II	III
2- 8	A	A	A	A	A	A	B
9- 15	A	A	A	A	A	B	C
16- 25	A	A	B	B	B	C	D
26- 50	A	B	B	C	C	D	E
51- 90	B	B	C	C	C	E	F
91- 150	B	B	C	D	D	F	G
151- 280	B	C	D	E	E	G	H
281- 500	B	C	D	E	F	H	J
501- 1.200	C	C	E	F	G	J	K
1.201- 3.200	C	D	E	G	H	K	L
3.201- 10.000	C	D	F	G	J	L	M
10.001- 35.000	C	D	F	H	K	M	N
35.001-150.000	D	E	G	J	L	N	P
150.001-500.000	D	E	G	J	M	P	Q
Mayor-500.000	D	E	H	K	N	Q	R

riores por un grupo de trabajo en representación de Estados Unidos, el Reino Unido y Canadá. Estas normas han sido progresivamente aceptadas en todo el mundo para la inspección por atributos. En la actualidad estos planes de muestreo se denominan en Estados Unidos MIL-STD-105D, en Europa ISO-2859 y en España son la norma UNE 66-020-73.

Estos planes están especialmente adaptados para ser usados en la inspección rutinaria de muchos lotes. Se caracterizan porque el tipo de inspección se va ajustando en función de la calidad de los lotes que se inspeccionan. Permiten muestreo simple, doble y mútiple. En el muestreo simple la decisión se toma en función de una única muestra. En el muestreo doble se toma una primera muestra pequeña y, según el número de defectos que se encuentren en ella, tomaremos la decisión de aceptar, tomar una segunda muestra o rechazar el lote. En los muestreos múltiples la idea es la misma, pero ahora, cuando los resultados no son concluyentes con la primera muestra, es posible tomar una segunda y así sucesivamente. Las tablas MIL-STD-105D permiten tomar hasta siete muestras, cada una de tamaño pequeño.

Vamos a exponer aquí únicamente las tablas y el procedimiento para el muestreo simple. Las tablas tienen en cuenta el AQL o valor de p_A, nivel de calidad aceptable, el nivel de inspección (que definiremos a continuación) y el tamaño del lote. Los planes están diseñados para que α sea aproximadamente 0,05, pero no tienen explícitamente en cuenta un valor de p_R y β, aunque para $p_R \triangleq 5p_A$ el riesgo β suele ser pequeño. Su funcionamiento es el siguiente:

1) Decidir el AQL, p_A.
2) Determinar el tipo de inspección, con los siguientes criterios:

Coste de inspección alto	Nivel I
Caso estándar	Nivel II
Coste de inspección bajo	Nivel III
Ensayos destructivos	Niveles S-1 a S-4.

3) Determinar el rigor de inspección. Éste puede ser:

   - *Riguroso:* cuando se sospecha que la calidad es peor que el AQL. Se pasará de la inspección normal a la rigurosa cuando dos de cinco lotes o partidas consecutivos sean rechazados en la inspección.
   - *Normal:* cuando se espere calidad similar al AQL. Deberá pasarse de riguroso a normal cuando se acepten cinco lotes seguidos con inspección rigurosa.
   - *Reducido:* cuando la calidad sea superior al AQL. Se pasará a ella al aceptar diez lotes consecutivos con inspección normal, y se volverá a la inspección normal al rechazar un lote (véase tabla 13.9).

4) Conocido el tamaño del lote y el tipo de inspección fijado en (2) utilizar la tabla 13.6 para obtener una letra-código de inspección.

Con la letra-código determinada en el paso anterior ir a la tabla de planes de inspección correspondiente al rigor definido en (3) y entrando en el AQL determinar el número de defectuosos para aceptar y rechazar (véanse las tablas 13.7, 13.8 y 13.9). Si en la casilla resultado de cruzar ambas variables hay una flecha, esto indica que no existe un plan con estas características y debemos escoger el más próximo siguiendo las flechas de casilla en casilla hasta que lleguemos a una casilla con un plan, especificado por el tamaño muestral y el número de defectuosos para aceptar y rechazar el lote.

Tabla 13.7 MIL II-A. Inspección normal

		Valores de p_A = AQL											
c	n	0,010	0,015	0,025	0,040	0,065	0,10	0,15	0,25	0,40	0,65	1,0	1,5
		Ac Re	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*
A	2	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
B	3												
C	5												
D	8												0 1
E	13											0 1	↕
F	20										0 1	↕	
G	32									0 1	↕	↓	1 2
H	50								0 1	↕	↓	1 2	2 3
J	80						↓	0 1	↕	↓	1 2	2 3	3 4
K	125						0 1	↕	↓	1 2	2 3	3 4	5 6
L	200					0 1	↕	↓	1 2	2 3	3 4	5 6	7 8
M	315			↓	0 1	↕	↓	1 2	2 3	3 4	5 6	7 8	10 11
N	500		↓	0 1	↕	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15
P	800		0 1	↕	↕	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22
Q	1.250	0 1	↕	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	↑
R	2.000	↑	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	|	

n = tamaño muestral; *c* = letra código obtenida de tabla 13.6; *Ac* = número de defectuosas para aceptar; *Re* = número para rechazar.
Los valores de AQL entre 0,01 y 10 representan % de elementos defectuosos o, si las unidades pueden tener más de un defecto, defectos por 100 unidades. A partir del valor 10 son sólo número de defectos por 100 unidades.

Valores de p_A = AQL

	2,5	4,0	6,5	10	15	25	40	65	100	150	250	400	650	1.000
	Ac Re	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*
	↓ 0 1	↓ 0 1 ↑	0 1 ↕	↓ 1 2	↓ 1 2 2 3	1 2 2 3 3 4	2 3 3 4 5 6	3 4 5 6 7 8	5 6 7 8 10 11	7 8 10 11 14 15	10 11 14 15 21 22	14 15 21 22 30 31	21 22 30 31 44 45	30 31 44 45 ↑
	↕ 1 2	↓ 1 2 2 3	1 2 2 3 3 4	2 3 3 4 5 6	3 4 5 6 7 8	5 6 7 8 10 11	7 8 10 11 14 15	10 11 14 15 21 22	14 15 21 22 ↑	21 22 30 31 ↑	30 31 44 45 ↑	44 45 ↑	↑	
	2 3 3 4 5 6	3 4 5 6 7 8	5 6 7 8 10 11	7 8 10 11 14 15	10 11 14 15 21 22	14 15 21 22 ↑	21 22 ↑	↑						
	7 8 10 11 14 15	10 11 14 15 21 22	14 15 21 22 ↑	21 22 ↑	↑									
	21 22 ↑	↑												

Tabla 13.8 MIL II-B. Inspección estricta

		Valores de p_A = AQL											
c	n	0,010	0,015	0,025	0,040	0,065	0,10	0,15	0,25	0,40	0,65	1,0	1,5
		Ac Re	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*
A	2	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
B	3												
C	5												
D	8												
E	13											↓	0 1
F	20										↓	0 1	↓
G	32									↓	0 1	↓	
H	50								↓	0 1	↓	↓	1 2
J	80							↓	0 1	↓	↓	1 2	2 3
K	125						↓	0 1	↓	↓	1 2	2 3	3 4
L	200					↓	0 1	↓	↓	1 2	2 3	3 4	5 6
M	315				↓	0 1	↓	↓	1 2	2 3	3 4	5 6	8 9
N	500			↓	0 1	↓	↓	1 2	2 3	3 4	5 6	8 9	12 13
P	800		↓	0 1	↓	↓	1 2	2 3	3 4	5 6	8 9	12 13	18 19
Q	1.250	↓	0 1	↓	↓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	↑
R	2.000	0 1		↓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	↑	
S	3.150			1 2									

n = tamaño muestral; *c* = letra código obtenida de tabla 13.6; *Ac* = número de defectuosas para aceptar; *Re* = número para rechazar.
Los valores de AQL entre 0,01 y 10 representan % de elementos defectuosos o, si las unidades pueden tener más de un defecto, defectos por 100 unidades. A partir del valor 10 son sólo número de defectos por 100 unidades.

Valores de p_A = AQL

	2,5	**4,0**	**6,5**	**10**	**15**	**25**	**40**	**65**	**100**	**150**	**250**	**400**	**650**	**1.000**
	Ac Re	***Ac Re***	***Ac Re***	***Ac Re***	***Ac Re***	***Ac Re***	***Ac Re***	***Ac Re***	***Ac Re***	***Ac Re***	***Ac Re***	***Ac Re***	***Ac Re***	***Ac Re***
	↓	↓ 0 1	↓ 0 1	↓	↓ 1 2	↓ 1 2 2 3	1 2 2 3 3 4	2 3 3 4 5 6	3 4 5 6 8 9	5 6 8 9 12 13	8 9 12 13 18 19	12 13 18 19 27 28	18 19 27 28 41 42	27 28 41 42 ↑
	0 1 ↓	↓ 1 2	↓ 1 2 2 3	1 2 2 3 3 4	2 3 3 4 5 6	3 4 5 6 8 9	5 6 8 9 12 13	8 9 12 13 18 19	12 13 18 19 ↑	18 19 27 28 ↑	27 28 41 42 ↑	41 42 ↑	↑	
	1 2 2 3 3 4	2 3 3 4 5 6	3 4 5 6 8 9	5 6 8 9 12 13	8 9 12 13 18 19	12 13 18 19 ↑	18 19 ↑	↑						
	5 6 8 9 12 13	8 9 12 13 18 19	12 13 18 19 ↑	18 19 ↑	↑									
	18 19 ↑													

Tabla 13.9 MIL II-C. Inspección reducida

c	n	Valores de p_A = AQL											
		0,010	0,015	0,025	0,040	0,065	0,10	0,15	0,25	0,40	0,65	1,0	1,5
		Ac Re	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*
A	2	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
B	2												
C	2												
D	3												0 1
E	5											0 1	↕
F	8									↓	0 1	↑	
G	13								↓	0 1	↕	↓	0 2
H	20							↓	0 1	↕	↓	0 2	1 3
J	32						↓	0 1	↑	↕	0 2	1 3	1 4
K	50					↓	0 1	↕	↓	0 2	1 3	1 4	2 5
L	80				↓	0 1	↕	↓	0 2	1 3	1 4	2 5	3 6
M	125			↓	0 1	↑	↕	0 2	1 3	1 4	2 5	3 6	5 8
N	200		↓	0 1	↕	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10
P	315		0 1	↕	↕	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13
Q	500	0 1	↑	↕	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	↓
R	800	↑	↑	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	↑	↓

n = tamaño muestral; *c* = letra código obtenida de tabla 13.6; *Ac* = número de defectuosas para aceptar; *Re* = número para rechazar.
Los valores de AQL entre 0,01 y 10 representan % de elementos defectuosos o, si las unidades pueden tener más de un defecto, defectos por 100 unidades. A partir del valor 10 son sólo número de defectos por 100 unidades.

Valores de p_A = AQL

	2,5	4,0	6,5	10	15	25	40	65	100	150	250	400	650	1.000
	Ac Re	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*	*Ac Re*
	↓	↓	0 1	↓	↓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31
	0 1	0 1	↑↓	0 2	0 2	1 3	2 4	3 5	5 6	7 8	10 11	14 15	21 22	30 31
		↑			1 3	1 4	2 5	3 6	5 8	7 10	10 13	14 17	21 24	↑
	↑↓	↓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	14 17	21 24	↑	
	0 2	0 2	1 3	1 4	2 5	3 6	5 8	7 10	↑	↑	↑	↑		
		1 3	1 4	2 5	3 6	5 8	7 10	10 13						
	1 3	1 4	2 5	3 6	5 8	7 10	10 13	↑						
	1 4	2 5	3 6	5 8	7 10	10 13	↑							
	2 5	3 6	5 8	7 10	10 13	↑								
	3 6	5 8	7 10	10 13	↑									
	5 8	7 10	10 13	↑										
	7 10	10 13	↑											
	10 13	↑												
	↑													

Ejemplo 13.8

Diseñar un plan de muestreo con el plan japonés y con Military Standard de las características siguientes: p_A = 1%; α = 0,05; p_R = 5%; β = 0,10, para lotes de 1.000 unidades.

El plan japonés se obtiene entrando con los datos dados en la tabla 13.5 y resulta ser: tomar 120 piezas, si $x > 3$ rechazar el lote.

Para determinar el plan MIL-Standard, supondremos nivel II (estándar) y rigor normal. Entonces, la tabla 13.6 proporciona la letra código *J* y la tabla 13.7 de inspección normal, entrando con AQL = 1% y código *J* resulta: tomar 80 piezas, si $x \geq 3$ rechazar el lote.

Para comprobar ambos planes calcularemos la curva característica. Utilizando la aproximación normal, se obtiene la tabla 13.10.

Tabla 13.10 Comparación de planes de muestreo. Φ(.) función de distribución de la normal estándar

Plan		Japonés	MIL-Standard
Curva característica		$P(x \leq 3{,}5)$	$P(x \leq 2{,}5)$
Distribución aprox. de x		$N(120p; \sqrt{120p})$	$N(80p; \sqrt{80p})$
ORDENADAS para p	1%	$\Phi(2{,}100) = 0{,}98$	$\Phi(1{,}90) = 0{,}97$
	2%	$\Phi(0{,}710) = 0{,}76$	$\Phi(0{,}71) = 0{,}76$
	3%	$\Phi(-0{,}052) = 0{,}48$	$\Phi(0{,}64) = 0{,}52$
	4%	$\Phi(-0{,}593) = 0{,}28$	$\Phi(-0{,}39) = 0{,}35$
	5%	$\Phi(-1{,}021) = 0{,}15$	$\Phi(-0{,}75) = 0{,}23$
	6%	$\Phi(-1{,}379) = 0{,}09$	$\Phi(-1{,}049) = 0{,}15$
	7%	$\Phi(-1{,}691) = 0{,}04$	$\Phi(-1{,}309) = 0{,}10$

Se observa que el riesgo de admitir lotes malos es mayor con Military Standard.

13.11.6 Planes de control rectificativo: Dodge-Romig

Los planes de control rectificativo se basan en que los lotes rechazados son inspeccionados al 100% y todos los elementos defectuosos se sustituyen por buenos. De esta manera se garantiza que la *calidad media de entrada en almacén* (AOQ, *Average Outgoing Quality*) será alta. En efecto, supongamos un lote de N unidades con D defectuosas, o proporción defectuosa $p_D = D/N$. Sea $\beta(p_D)$ la probabilidad de que un plan dado acepte lotes con dicha calidad. Entonces, al realizar el muestreo puede ocurrir:

El lote es:	Probabilidad	Proporción defectuosa que entra en el almacén
Aceptado	$\beta(p_D)$	p_D
Rechazado	$1 - \beta(p_D)$	0

ya que los lotes rechazados se inspeccionan al 100%. Por tanto, la calidad promedio en el almacén será:

$$\text{AOQ} = \beta(p_D)p_D + [1 - \beta(p_D)] \cdot 0 = \beta(p_D)p_D$$

El valor AOQ depende pues de p, y se calcula multiplicando la curva característica por p, con lo que obtenemos la curva AOQ(p), que representa la calidad de salida. Su forma se presenta en la figura 13.22.

Figura 13.22 Curva de calidad promedio en el almacén

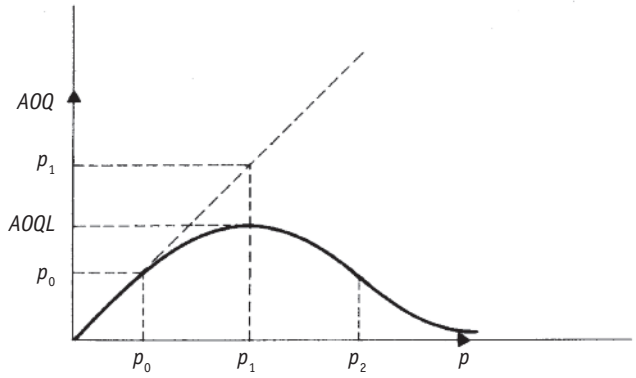


Tabla 13.11 Ejemplo de tablas de tolerancia de lote de muestreo simple, Dodge-Romig

Tamaño del lote	Promedio del proceso 0 a 0,04%			Promedio del proceso 0,05 a 0,40%			Promedio del proceso 0,41 a 0,80%		
	n	*c*	AOQL %	*n*	*c*	AOQL %	*n*	*c*	AOQL %
1-35	Todos	0	0	Todos	0	0	Todos	0	0
36-50	34	0	0,35	34	0	0,35	34	0	0,35
51-100	44	0	0,47	44	0	0,47	44	0	0,47
101-200	50	0	0,55	50	0	0,55	50	0	0,55
201-300	55	0	0,57	55	0	0,57	85	1	0,71
301-400	55	0	0,58	55	0	0,58	90	1	0,72
401-500	55	0	0,60	55	0	0,60	90	1	0,77
501-600	55	0	0,61	95	1	0,76	125	2	0,87
601-800	55	0	0,62	95	1	0,78	125	2	0,93
801-1.000	55	0	0,63	95	1	0,80	130	2	0,92
1.001-2.000	55	0	0,65	95	1	0,84	165	3	1,1
2.001-3.000	95	1	0,86	130	2	1,0	165	3	1,1
3.001-4.000	95	1	0,86	130	2	1,0	195	4	1,2
4.001-5.000	95	1	0,87	130	2	1,0	195	4	1,2
5.001-7.000	95	1	0,87	130	2	1,0	200	4	1,2
7.001-10.000	95	1	0,88	130	2	1,1	230	5	1,4
10.001-20.000	95	1	0,88	165	3	1,2	265	6	1,4
20.001-50.000	95	1	0,88	165	3	1,2	295	7	1,5
50.001-100.000	95	1	0,88	200	4	1,3	325	8	1,6

(c = número máximo de defectos que pueden aceptarse, nivel de calidad rechazable = 4%, $\beta = 0{,}10$).

Cuando la calidad de entrada p es muy buena (punto p_0), la calidad en el almacén será buena y próxima a p_0, porque todos los lotes se aceptarán. A medida que la proporción defectuosa aumenta, lo hará también AOQ, aunque en menor proporción, ya que los lotes comenzarán a ser rechazados e inspeccionados al 100%. La curva llega a un máximo que se denomina *lí-*

Promedio del proceso 0,81 a 1,20%			Promedio del proceso 1,21 a 1,60%			Promedio del proceso 1,61 a 2,00%		
n	***c***	AOQL %	***n***	***c***	AOQL %	***n***	***c***	AOQL %
Todos	0	0	Todos	0	0	Todos	0	0
34	0	0,35	34	0	0,35	34	0	0,35
44	0	0,47	44	0	0,47	44	0	0,47
50	0	0,55	50	0	0,55	50	0	0,55
85	1	0,71	85	1	0,71	85	1	0,71
120	2	0,80	120	2	0,80	145	3	0,86
120	2	0,87	150	3	0,91	150	3	0,91
125	2	0,87	155	3	0,93	185	4	0,95
160	3	0,97	190	4	1,0	220	5	1,0
165	3	0,98	220	5	1,1	255	6	1,1
195	4	1,2	255	6	1,3	315	8	1,4
230	5	1,3	320	8	1,4	405	11	1,6
260	6	1,4	350	9	1,5	465	13	1,6
290	7	1,4	380	10	1,6	520	15	1,7
290	7	1,5	410	11	1,7	575	17	1,9
325	8	1,5	440	12	1,7	645	19	1,9
355	9	1,6	500	14	1,8	730	22	2,0
380	10	1,7	590	17	2,0	870	26	2,1
410	11	1,8	620	18	2,0	925	29	2,2

mite de la calidad media de entrada (AOQL, *Average Outgoing Quality Limit*):

$$\text{AOQL} = \max_{p} \text{AOQ}(p)$$

para decrecer a partir de ahí. Lotes muy malos (punto p_2) serán casi siempre rechazados y por tanto conducirán a una alta calidad en el almacén.

Las tablas ML-STD-105D proporcionan el AOQL para muchos planes de muestreo. Sin embargo, no están diseñadas para definir el plan en función del AOQL. Unos planes de muestreo diseñados con este objetivo son planes de muestreo de Dodge-Romig, que pueden utilizarse de dos formas:

1) Fijando N, tamaño del lote, el promedio del proceso (AQL) y el nivel de calidad rechazable. Entonces proporcionan n, tamaño muestral, c (número máximo aceptable de defectos) y el AOQL del plan con control rectificativo. Las tablas están construidas con β, riesgo del consumidor, igual a 0,10.
2) Fijando N, el AQL y el AOQL. Se obtiene n, c y el nivel de calidad rechazable.

La tabla 13.11 presenta un ejemplo de estos planes. Está diseñada para $p_R = 0{,}04$, $\beta = 0{,}10$ y proporciona el AOQL del plan resultante.

Ejercicios 13.2

13.2.1. Calcular un plan de muestreo con Military-Standard para lotes de tamaño 1.500, inspección normal y AQL = 1,5% de defectuosos.

13.2.2. En un control de recepción se admite que para $p = 0{,}01$ deben aceptarse lotes con $\alpha \simeq 0{,}05$ y que cuando $p = 0{,}04$ deben rechazarse con alta probabilidad; los lotes son de 5.000 unidades y, en el pasado, la calidad suministrada ha sido próxima al 1%. Utilizar:

a) Military-Standard.
b) Dodge-Romig.
c) Compararlos.

13.2.3. En la recepción de un lote de 4.000 unidades el comprador puede optar entre un muestreo con las tablas Dodge-Romig o Military-Standard. Si AQL = 1%, NCR = 4%, $\beta = 0{,}1$, comparar ambos planes.

13.2.4. En un control de recepción por variables se toma una muestra de tamaño n y se acepta si $\bar{x} \leq c$. Determinar n y c con la condición de que cuando $\mu = 200$ la probabilidad de aceptar sea 0,95 y cuando $\mu = 210$ esta probabilidad sea de 0,10. Se conoce que $\sigma = 10$.

13.2.5. En un control de recepción de lotes de 2.000 piezas se toman $n = 100$, aceptando el lote si $x < 3$ defectuosas. Dibujar la curva OC y la curva de calidad de salida promedio utilizando la aproximación de Poisson.

13.2.6. Una empresa compra mucho material con control de recepción $n = 5$, $c = 0$.

a) Dibujar la curva OC con la distribución binomial.
b) Dibujar la curva AOQ.
c) Estimar el AOQL.
d) Estimar el AQL si $\alpha = 0{,}05$.
e) Estimar NCR si $\beta = 0{,}10$.

13.2.7. Se define el siguiente plan de muestreo: se toman 50 piezas, si en la muestra hay una o ninguna defectuosa se acepta, si hay más de cuatro se rechaza, si hay 2 o 3 se toma otra segunda muestra de 50 piezas y si en el total de las dos muestras hay 4 o más defectuosas se rechaza la partida, aceptándola en otro caso.
Se pide:

a) Probabilidad de aceptar una partida con el 2% de defectos.
b) Determinar la esperanza matemática del número de piezas necesario para tomar una decisión.

13.2.8. Se utiliza el plan doble siguiente: tomar cinco piezas, si todas están bien aceptar, en caso contrario tomar otras 5, si el total de defectos en las diez es uno o menos, aceptar; si no rechazar.
Se pide:

a) Construir la curva OC.
b) Construir la curva AOQ.
c) Calcular el AOQL.
d) Estimar el AQL si $\alpha = 0{,}05$.
e) Estimar el NCR si $\beta = 0{,}10$.

13.12 Resumen del capítulo

Controlar un proceso es comprobar que se encuentra en estado de control: la distribución de las características de interés es constante. El control de procesos se realiza tomando muestras a lo largo del tiempo, midiendo estas características y comprobando mediante un contraste de hipótesis que la distribución no varía.

Los gráficos de control son las herramientas técnicas básicas del control estadístico de procesos. Incluyen una línea central al nivel de la media esperable del proceso y dos líneas de control a tres desviaciones típicas de la media. Cuando las características de la muestra están fuera de los límites de control, concluimos que ha cambiado la distribución y buscamos las causas para eliminarlas.

El control de recepción se aplica para verificar que una partida de productos (materia prima o producto final) cumple las especificaciones establecidas. Consiste de nuevo en un contraste de hipótesis que se construye teniendo en cuenta los dos posibles tipos de error: aceptar un lote malo (riesgo del receptor) y rechazar un lote bueno (riesgo del suministrador).

Los programas estadísticos habituales incluyen la posibilidad de construir gráficos de control y estudios de capacidad.

13.13 Lecturas recomendadas

El libro de Bowker y Lieberman (1981) contiene una buena exposición del control de calidad tanto en proceso (capítulo 12) como en recepción (capítulos 13 y 14). Grant y Leavenworth (1986) es un clásico del control estadístico de calidad, y Juran *et al.* (1983) incluye una enciclopédica recopilación de trabajos, muchos de ellos relacionados con la gerencia de calidad. Prat y otros (2004) es una buena presentación de métodos de control y mejora de la calidad. Otros libros de interés son: Braverman (1981), pensado para personas sin conocimientos previos de estadística; Charbonneau y Webster (1983), tratamiento ingenieril del control de calidad; Hald (1981), centrado en el estudio matemático del control de recepción por atributos; Ishikawa (1990), que muestra cómo utilizar técnicas estadísticas simples para mejorar la calidad y la productividad; Montgomery (2008), que incluye cómo diseñar con criterios económicos gráficos de control; Ott (1975), que contiene muchos estudios de casos, y Taguchi (1981), Taguchi *et al.* (2004) y Ross (1995), que presentan el enfoque japonés al control estadístico de calidad.

El lector interesado en una visión más amplia de la calidad debe consultar los textos de Ledolter y Burrell (1999), Deming (2000), Juran (1983, 1995) y Tiao *et al.* (2005). Para los nuevos enfoques seis sigma, Pyzdek (1999), Breyfogle (2003) y Harry y Schroder (2006).

Apéndice 13A: Cálculo de gráficos de control

Medias

En el apéndice 5C se estudió la función de densidad gamma, que se reduce a la χ^2 si $\lambda = \frac{1}{2}$, $n = 2r$. Por tanto, la distribución χ^2 con n grados de libertad tiene como función de densidad:

$$f(x) = \frac{1}{\Gamma(n/2)} (\tfrac{1}{2})^{n/2}\, x^{(n/2)-1}\, e^{-x/2} \qquad x > 0$$

donde $\Gamma(n/2)$ es la función gamma definida para z entero por $\Gamma(z) = (z-1)!$ y $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. Como ns^2/σ^2 es una χ^2 con $n-1$ grados de libertad, la distribución de s^2 se obtendrá de:

$$f(s^2) = f(\chi^2_{n-1}) \left| \frac{d\chi^2}{ds^2} \right| = f\left(\frac{ns^2}{\sigma^2} \right) \frac{n}{\sigma^2}$$

resultando la función de densidad de s^2:

$$f(s^2) = \frac{n/\sigma^2}{2^{(n-1)/2}\, \Gamma\left(\dfrac{n-1}{2} \right)} \left(\frac{ny}{\sigma^2} \right)^{(n-3)/2} e^{-ny/2\sigma^2} \qquad y > 0$$

Para obtener la distribución de s utilizamos de nuevo la fórmula de cambio de variable. Llamando $y = s^2$:

$$f(s) = f(y \left| \frac{dy}{ds} \right| = f(s^2) 2s$$

y operando se obtiene:

$$f(s) = \frac{2 \left(\dfrac{n}{2} \right)^{\left(\frac{n-1}{2} \right)}}{\sigma^{n-1}\, \Gamma\left(\dfrac{n-1}{2} \right)} s^{n-2}\, e^{-ns^2/2\sigma^2} \qquad s > 0 \qquad \text{(13A.1)}$$

Integrando en (13A.1) y haciendo el cambio $ns^2/2\sigma^2 = z$ la integral se convierte en la función gamma y se obtiene finalmente:

$$E[s] = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \sqrt{\frac{2}{n}}\, \sigma = c_2 \sigma \qquad (13A.2)$$

y dando valores a n se obtiene la tabla 13.1. Comparemos esta expresión exacta con la aproximación obtenida en (7.17) para la media de $\hat{s}$:

$$E[\hat{s}] = \frac{4n-5}{4n-4}\, \sigma$$

Utilizando esta aproximación, la esperanza de s será:

$$E[s] = \sqrt{\frac{n-1}{n}}\, E[\hat{s}] = \sqrt{\frac{n-1}{n}} \left(\frac{4n-5}{4n-4} \right) \sigma \qquad (13A.3)$$

por ejemplo, para $n = 10$ la expresión (13A.2) conduce a 0,9227 y la (13A.3) a 0,9223. La aproximación (13A.3) es muy buena para pequeños tamaños de n y mucho más fácil de usar que (13A.2).

Desviaciones típicas

Para calcular la desviación típica de la distribución de s partiremos de

$$DT(s) = \sqrt{E(s^2) - E^2(s)}.$$

Como $E(s^2) = \sigma^2(n-1)/n$ y $E(s) = c_2\sigma$, tenemos que

$$DT(s) = \sigma \sqrt{(n-1)/n - c_2^2}$$

y estimando σ por $\bar{s}/c_2$, los límites superior (LS) e inferior (LI) del intervalo serán:

$$LS = \bar{s} + 3\, \frac{\bar{s}\sqrt{(n-1)/n - c_2^2}}{c_2} = \bar{s} B_4$$

$$LI = \bar{s} - 3\, \frac{\bar{s}\sqrt{(n-1)/n - c_2^2}}{c_2} = \bar{s} B_3$$

donde $B_4 = (1 + 3\sqrt{(n-1)/nc_2^2 - 1})$ y $B_3 = (1 - 3\sqrt{(n-1)/nc_2^2 - 1})$.

El intervalo es análogo si utilizamos $\hat{s}$. Entonces $E(\hat{s}^2) = \sigma^2$ y $E(\hat{s}) = c_2\sigma\sqrt{n}/\sqrt{n-1} = c_4\sigma$, donde

$$c_4 = c_2\sqrt{n}/\sqrt{n-1}$$

La desviación típica de $\hat{s}$ será

$$DT(\hat{s}) = \sigma\sqrt{1 - c_4^2}$$

y el intervalo, llamando ahora $\hat{\bar{s}} = \Sigma\, \hat{s}_i/k$ se calculará como

$$LS = \hat{\bar{s}} + 3\,\frac{\hat{\bar{s}}\sqrt{1 - c_4^2}}{c_4} = \hat{\bar{s}}(1 + 3\,\sqrt{1/c_4^2 - 1}) = B_4\hat{\bar{s}}$$

ya que por la relación entre c_2 y c_4 ambos conducen a las mismas constantes.

Tablas

Explicación de las tablas

Tabla 1: Números aleatorios

Entrando por cualquier fila y columna se obtienen números aleatorios de cualquier tamaño. Para utilizarlos en los ejercicios del texto, se recomienda tomar únicamente tres dígitos, añadiendo la coma decimal. Por ejemplo, descendiendo en vertical el primer bloque de números se obtiene: 0,593; 0,995; 0,103, etc.

Tabla 2: Probabilidades binomiales

Proporciona las probabilidades acumuladas de la distribución binomial mediante la fórmula:

$$P(x \leq k) = \sum_{j=0}^{k} \binom{n}{j} p^j (1-p)^{n-j}$$

Los valores de n y k se especifican verticalmente y p horizontalmente.

Si $p > 0{,}5$ se aplica que:

$$P(x = k \mid \text{dadas } n \text{ y } p) = P(x = n - k \mid \text{dadas } n \text{ y } p' = 1 - p)$$

Ejemplo 2.1

Calcular la probabilidad de más de 2 éxitos en cinco pruebas si $P = 0{,}8$.

$$P(x > 2 \mid n = 5, p = 0{,}8) = P(X < 3 \mid n = 5, P = 0{,}2)$$

$$\sum_{i=3}^{5} P(X = i \mid n = 5, P = 0{,}8) = \sum_{0}^{2} P(X = i \mid n = 5, P = 0{,}2) = 0{,}9421$$

Tabla 3: Probabilidades de Poisson

Proporciona probabilidades acumuladas de la distribución de Poisson, mediante:

$$P(x \leqslant k) = \sum_{x=0}^{k} e^{-\lambda} \frac{\lambda^x}{x!}$$

Ejemplo 3.1

Calcular la probabilidad de que $x = 3$ en una distribución de Poisson con $\lambda = 2$

$$P(x = 3) = P(x \leqslant 3) - P(x \leqslant 2) = 0{,}857 - 0{,}677 = 0{,}18$$

Tabla 4: Distribución normal

Esta tabla proporciona valores de la función de distribución para la distribución normal estandarizada (con media cero y varianza unidad).

La tabla contiene únicamente los datos para valores positivos de *z,* ya que, por simetría:

$$F(-|z_0|) = P(z \leqslant -|z_0|) = P(z \geqslant |z_0|) = 1 - F(z_0)$$

Para obtener las probabilidades correspondientes a una normal general (μ, σ) estandarizamos la variable antes de entrar en las tablas.

Ejemplo 4.1

Obtener la probabilidad del suceso $5 < x \leqslant 10$ si x es normal con media 7 y desviación típica 2.

$$P(5 < x \leqslant 10) = P\left(\frac{5-7}{2} < \frac{x-7}{2} \leqslant \frac{10-7}{2}\right) = P(-1 < z \leqslant 1{,}5)$$

donde z es ahora una variable normal (0,1).

$$P(-1 < z \leqslant 1{,}5) = P(z \leqslant 1{,}5) - P(z \leqslant -1)$$

utilizando la simetría de la curva $N(0,1)$ alrededor de su media, cero, tendremos que:

$$P(z \leqslant -1) = P(z \geqslant 1) = 1 - F(1)$$

Con las tablas obtenemos:

$$F(1{,}5) = 0{,}93319$$
$$F(-1) = 1 - F(1) = 1 - 0{,}8413 = 0{,}1587$$

Por lo tanto:

$$P(-1 < z \leqslant 1{,}5) = 0{,}93319 - 0{,}1587 = 0{,}774619$$

Ejemplo 4.2

Encontrar un valor x_1 tal que $P(3 < x \leqslant x_1) = 0{,}5$, si x es normal (7,2).

$$P(3 < x \leqslant x_1) = P\left(\frac{3-7}{2} < \frac{x-7}{2} \leqslant \frac{x_1-7}{2}\right) =$$
$$= P\left(-2 < z \leqslant \frac{x_1-7}{2}\right) =$$
$$= P\left(z \leqslant \frac{x_1-7}{2}\right) - P(z \leqslant -2) = 0{,}5$$

Como: $P(z \leqslant -2) = P(z \geqslant 2) = 1 - F(2) = 0{,}0288$

$$P\left(z \leqslant \frac{x_1-7}{2}\right) = 0{,}5 + 0{,}0228 = 0{,}5228$$

Tanteando en las tablas:

$$P(z \leqslant 0{,}05) = 0{,}5199$$
$$P(z \leqslant 0{,}06) = 0{,}5239$$

El valor buscado está situado entre ambos. Interpolando linealmente:

$$\frac{x_1 - 7}{2} = 0{,}05 + \frac{0{,}5228 - 0{,}5199}{0{,}5329 - 0{,}5199} \cdot 0{,}01 = 0{,}0572$$

$$x_1 = 7 + 2\,(0{,}0572) = 7{,}1144$$

Tabla 5: Distribución *t*

Proporciona los percentiles de la distribución *t* de Student, en función del número de grados de libertad (*k*). Recordando que esta distribución es simétrica con media cero, podemos obtener los percentiles complementarios a los que aparecen en las tablas.

Ejemplo 5.1

Calcular dos valores tales que $P(t_a < t \leqslant t_b) = 0{,}95$ para una distribución con 10 grados de libertad, tomando un intervalo simétrico.

Si $P(t \leqslant t_a) = P(t \geqslant t_b)$ para que sea simétrico,

$$P(t \leqslant t_a) = P(t \geqslant t_b) = 0{,}025$$

Por lo tanto $t_{0{,}975}$ para 10 grados de libertad es 2,23 y

$$t_b = 2{,}23$$

$$t_a = -2{,}23$$

Para construir un intervalo no simétrico, tomando $t_a = -\infty$:

$$P(t \leqslant t_b) = 0{,}95$$

Mirando en tablas $t_{0{,}975}$ obtenemos:

$$t_b = 1{,}81$$

Ejemplo 5.2

Obtener $P(t \leqslant 2{,}2)$ para una *t* con 15 grados de libertad.

De las tablas obtenemos:

$$P(t \leqslant 2{,}13) = 0{,}975$$

$$P(t \leqslant 2{,}6) = 0{,}990$$

e interpolando linealmente:

$$P(t \leqslant 2{,}2) = 0{,}975 + \frac{0{,}990 - 0{,}975}{2{,}6 - 2{,}13}(2{,}2 - 2{,}13) = 0{,}977$$

Tabla 6: Distribución χ^2

Proporciona los percentiles para una χ^2 con k grados de libertad. Su utilización es similar a la distribución t pero con la diferencia de que la distribución chi-cuadrado no es simétrica.

Para $k > 100$ se utiliza la aproximación de que $\sqrt{2\chi^2_k}$ se distribuye de forma asintóticamente normal con parámetros ($\sqrt{2k-1}$, 1).

Ejemplo 6.1

Calcular dos valores χ^2_a y χ^2_b que verifiquen $P(\chi^2_a < \chi^2 \leqslant \chi^2_b) = 0{,}95$ en una distribución con $K = 10$.

Tomando un intervalo simétrico en probabilidad:

$$P(\chi^2 \leqslant \chi^2_a) = 0{,}025 \Rightarrow \chi^2_{0{,}025} = \chi^2_a = 3{,}25$$
$$P(\chi^2 \geqslant \chi^2_b) = 0{,}025 \Rightarrow \chi^2_{0{,}975} = \chi^2_b = 18{,}3$$
$$P(3{,}25 < \chi^2 \leqslant 18{,}3) = 0{,}95$$

Ejemplo 6.2

Calcular $P(\chi^2 \leqslant 16)$ para 12 grados de libertad.

$$P(\chi^2 \leqslant 18{,}5) = 0{,}9$$
$$P(\chi^2 \leqslant 14{,}8) = 0{,}75$$

Interpolando

$$P(\chi^2 \leqslant 16) = 0{,}75 + \frac{0{,}9 - 0{,}75}{18{,}5 - 14{,}8}(16 - 14{,}8) = 0{,}7986$$

Ejemplo 6.3

Calcular $P(\chi^2_{200} > 200)$. Como $\sqrt{2\chi^2_{200}} \sim N(19{,}97;\ 1)$

$$P(\chi^2_{200} > 200) = P(\sqrt{2\chi^2_{200}} > \sqrt{400}) =$$
$$= P(z > 20 - 19{,}97) = P(z > 0{,}03) = 1 - 0{,}512 = 0{,}488$$

Tabla 7: Distribución *F*

Proporciona los percentiles 0,95 (tabla 7a) y 0,99 (tabla 7b) de la distribución $F(n_1, n_2)$ siendo n_1 el número de grados de libertad del numerador y n_2 los grados de libertad del denominador.

Tablas 8, 9, 10, 11 y 12

La explicación de estas tablas y de su manejo práctico se encuentra en el capítulo 12.

Tabla 1 Números aleatorios

	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-94	95-99
00	59391	58030	52098	82718	87024	82848	04190	96574	90464	29065
01	99567	76364	77204	04615	27062	96621	43918	01896	83991	51141
02	10363	97518	51400	25670	98342	61891	27101	37855	06235	33316
03	86859	19558	64432	16706	99612	59798	32803	67708	15297	28612
04	11258	24591	36863	55368	31721	94335	34936	02566	80972	08188
05	95068	88628	35911	14530	33020	80428	39936	31855	34334	64865
06	54463	47237	73800	91017	36239	71824	83671	39892	60518	37092
07	16874	62677	57412	13215	31389	62233	80827	73917	82802	84420
08	92494	63157	76593	91316	03505	72389	96363	52887	01087	66091
09	15669	56689	35682	40844	53256	81872	35213	09840	34471	74441
10	99116	75486	84989	23476	52967	67104	39495	39100	17217	74073
11	15696	10703	65178	90637	63110	17622	53988	71087	84148	11670
12	97720	15369	51269	69620	03388	13699	33423	67453	43269	56720
13	11666	13841	71681	98000	35979	39719	81899	07449	47985	46967
14	71628	73130	78783	75691	41632	09847	61547	18707	85489	69944
15	40501	51089	99943	91843	41995	88931	73631	69361	05375	15417
16	22518	55576	98215	82068	10798	86211	36584	67466	69373	40054
17	75112	30485	62173	02132	14878	92879	22281	16783	86352	00077
18	80327	02671	98191	84342	90813	49268	95441	15496	20168	09271
19	60251	45548	02146	05597	48228	81366	34598	72856	66762	17002
20	57430	82270	10421	00540	43648	75888	66049	21511	47676	33444
21	73528	39559	34434	88596	54086	71693	43132	14414	79949	85193
22	25991	65959	70769	64721	86413	33475	42740	06175	82758	66248
23	78388	16638	09134	59980	63806	48472	39318	35434	24057	74739
24	12477	09965	96657	57994	59439	76330	24596	77515	09577	91871
25	83266	32883	42451	15579	38155	29793	40914	65990	16255	17777
26	76970	80876	10237	39515	79152	74798	39357	09054	73579	92359
27	37074	65198	44785	68624	98336	84481	97610	78735	46703	98265
28	83712	06514	30101	78295	54656	85417	43189	60048	72781	72606
29	20287	56862	69727	94443	64936	08366	27227	05158	50326	59566
30	74261	32592	86538	27041	65172	85532	07571	80609	39285	65340
31	64081	49863	08478	96001	18888	14810	70545	89755	59064	07210
32	05617	75818	47750	67814	29575	10526	66192	44464	27058	40467
33	26793	74951	95466	74307	13330	42664	85515	20632	05497	33625
34	65988	72850	48737	54719	52056	01596	03845	35067	03134	70322
35	27366	42271	44300	73399	21105	03280	73457	43093	05192	48657
36	56760	10909	98147	34736	33863	95256	12731	66598	50771	83665
37	72880	43338	93643	58904	59543	23943	11231	83268	65938	81581
38	77888	38100	03062	58103	47961	83841	25878	23746	55903	44115
39	28440	07819	21580	51459	47971	29882	13990	29226	23608	15873
40	63525	94441	77033	12147	51054	49955	58312	76923	96071	05813
41	47606	93410	16359	89033	89696	47231	64498	31776	05383	39902
42	52669	45030	96279	14709	52372	87832	02735	50803	72744	88208
43	16738	60159	07425	62369	07515	82721	37875	71153	21315	00132
44	59348	11695	45751	15865	74739	05572	32688	20271	65128	14551
45	12900	71775	29845	60774	94924	21810	38636	33717	67598	82521
46	75086	23537	49939	33595	13484	97588	28617	17979	70749	35234
47	99495	51434	29181	09993	38190	42553	68922	52125	91077	40197
48	26075	31671	45386	36583	93459	48599	52022	41330	60651	91321
49	13636	93596	23377	51133	95126	61496	42474	45141	46660	42338

Tabla 1 Números aleatorios *(continuación)*

	00-04	05-09	10-14	15-l9	20-24	25-29	30-34	35-39	40-44	45-49
50	64249	63664	39652	40646	97306	31741	07294	84149	46797	82487
51	26538	44249	04050	48174	65570	44072	40192	51153	11397	58212
52	05845	00512	78630	55328	18116	69296	91705	86224	29503	57071
53	74897	68373	67359	51014	33510	83048	17056	72506	82949	54600
54	20872	54570	35017	88132	25730	22626	86723	91691	13191	77212
55	31432	96156	89177	75541	81355	24480	77243	76690	42507	84362
56	66890	61505	01240	00660	05873	13568	76082	79172	57913	93448
57	41894	57790	79970	33106	86904	48119	52503	24130	72824	21627
58	11303	87118	81471	52936	08555	28420	49416	44448	04269	27029
59	54374	57325	16947	45356	78371	10563	97191	53798	12693	27928
60	64852	34421	61046	90849	13966	39810	42699	21753	76192	10508
61	16309	20384	09491	91588	97720	89846	30376	76970	23063	35894
62	42587	37065	24526	72602	57589	98131	37292	05967	26002	51945
63	40177	98590	97161	41682	84533	67588	62036	49967	01990	72308
64	82309	76128	93965	26743	24141	04838	40254	26065	07938	76236
65	79788	68243	59732	04257	27084	14743	17520	95401	55811	76099
66	40538	79000	89559	25026	42274	23489	34502	75508	06059	86682
67	64016	73598	18609	73150	62463	33102	45205	87440	96767	67042
68	49767	12691	17903	93871	99721	79109	09425	26904	07419	76013
69	76974	55108	29795	08404	82684	00497	51126	79935	57450	55671
70	23854	08480	85983	96025	50117	64610	99425	62291	86943	21541
71	68973	70551	25098	78033	98573	79848	31778	29555	61446	23037
72	36444	93600	65350	14971	25325	00427	52073	64280	18847	24768
73	03003	87800	07391	11594	21196	00781	32550	57158	58887	73041
74	17540	26188	36647	78386	04558	61463	57842	90382	77019	24210
75	38916	55809	47982	41968	69760	79422	80154	91486	19180	15100
76	64288	19843	69122	42502	48508	28820	59933	72998	99942	10515
77	86809	51564	38040	39418	49915	19000	58050	16899	79952	57849
78	99800	99566	14742	05028	30033	94889	53381	23656	75787	59223
79	92345	31890	95712	08279	91794	94068	49337	88674	35355	12267
80	90363	65162	32245	82279	79256	80834	06088	99462	56705	06118
81	64437	32242	48431	04835	39070	59702	31508	60935	22390	52246
82	91714	53662	28373	34333	55791	74758	51144	18827	10704	76803
83	20902	17646	31391	31459	33315	03444	55743	74701	58851	27427
84	12217	86007	70371	52281	14510	76094	96579	54853	78339	20839
85	45177	02863	42307	53571	22532	74921	17735	42201	80540	54721
86	28325	90814	08804	52746	47913	54577	47525	77705	95330	21866
87	29019	28776	56116	54791	64604	08815	46049	71186	34650	14994
88	84979	81353	56219	67062	26146	82567	33122	14124	46240	92973
89	50371	26347	48513	63915	11158	25563	91915	18431	92978	11591
90	53422	06825	69711	67950	64716	18003	49581	45378	99878	61130
91	67453	35651	89316	41620	32048	70225	47597	33137	31443	51445
92	07294	85353	74819	23445	68237	07202	99515	62282	53809	26685
93	79544	00302	45338	16015	66613	88968	14595	63836	77716	79596
94	64144	85442	82060	46471	24162	39500	87351	36637	42833	71875
95	90919	11883	58318	00042	52402	28210	34075	33272	00840	73268
96	06670	57353	86275	92276	77591	46924	60839	55437	03183	13191
97	36634	93976	52062	83678	41256	60948	18685	48992	19462	96062
98	75101	72891	85745	67106	26010	62107	60885	37503	55461	71213
99	05112	71222	72654	51583	05228	62056	57390	42746	39272	96659

Tabla 2 Probabilidades binomiales acumuladas

		P											
n	**k**	**0,01**	**0,05**	**0,10**	**0,15**	**0,20**	**0,25**	**0,30**	**1/3**	**0,35**	**0,40**	**0,45**	**0,50**
2	**0**	0,9801	0,9025	0,8100	0,7225	0,6400	0,5625	0,4900	0,4444	0,4225	0,3600	0,3025	0,2500
	1	0,9999	0,9975	0,9900	0,9775	0,9600	0,9375	0,9100	0,8889	0,8775	0,8400	0,7975	0,7500
	2	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
3	**0**	0,9703	0,8574	0,7290	0,6141	0,5120	0,4219	0,3430	0,2963	0,2746	0,2160	0,1664	0,1250
	1	0,9997	0,9928	0,9720	0,9392	0,8960	0,8438	0,7840	0,7407	0,7182	0,6480	0,5748	0,5000
	2	1,0000	0,9999	0,9990	0,9966	0,9920	0,9844	0,9730	0,9630	0,9571	0,9360	0,9089	0,8750
	3		1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
4	**0**	0,9606	0,8145	0,6561	0,5220	0,4096	0,3164	0,2401	0,1975	0,1785	0,1296	0,0915	0,0625
	1	0,9994	0,9860	0,9477	0,8905	0,8192	0,7383	0,6517	0,5926	0,5630	0,4752	0,3910	0,3125
	2	1,0000	0,9995	0,9963	0,9880	0,9728	0,9492	0,9163	0,8889	0,8735	0,8208	0,7585	0,6875
	3		1,0000	0,9999	0,9995	0,9984	0,9961	0,9919	0,9876	0,9850	0,9744	0,9590	0,9375
	4			1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
5	**0**	0,9510	0,7738	0,5905	0,4437	0,3277	0,2373	0,1681	0,1317	0,1160	0,0778	0,0503	0,0312
	1	0,9990	0,9774	0,9185	0,8352	0,7373	0,6328	0,5282	0,4609	0,4284	0,3370	0,2562	0,1875
	2	1,0000	0,9988	0,9914	0,9734	0,9421	0,8965	0,8369	0,7901	0,7648	0,6826	0,5931	0,5000
	3		1,0000	0,9995	0,9978	0,9933	0,9844	0,9692	0,9547	0,9460	0,9130	0,8688	0,8125
	4			1,0000	0,9999	0,9997	0,9990	0,9976	0,9959	0,9947	0,9898	0,9815	0,9688
	5				1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
6	**0**	0,9415	0,7351	0,5314	0,3771	0,2621	0,1780	0,1176	0,0878	0,0754	0,0467	0,0277	0,0156
	1	0,9986	0,9672	0,8857	0,7765	0,6553	0,5339	0,4202	0,3512	0,3191	0,2333	0,1636	0,1094
	2	1,0000	0,9978	0,9842	0,9527	0,9011	0,8306	0,7443	0,6804	0,6471	0,5443	0,4415	0,3438
	3		0,9999	0,9987	0,9941	0,9830	0,9624	0,9295	0,8999	0,8826	0,8208	0,7447	0,6562
	4		1,0000	0,9999	0,9996	0,9984	0,9954	0,9891	0,9822	0,9777	0,9590	0,9308	0,8906
	5			1,0000	1,0000	0,9999	0,9998	0,9993	0,9986	0,9982	0,9959	0,9917	0,9844
	6					1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
7	**0**	0,9321	0,6983	0,4783	0,3206	0,2097	0,1335	0,0824	0,0585	0,0490	0,0280	0,0152	0,0078
	1	0,9980	0,9556	0,8503	0,7166	0,5767	0,4449	0,3294	0,2634	0,2338	0,1586	0,1024	0,0625
	2	1,0000	0,9962	0,9743	0,9262	0,8520	0,7564	0,6471	0,5706	0,5323	0,4199	0,3164	0,2266
	3		0,9998	0,9973	0,9879	0,9667	0,9294	0,8740	0,8267	0,8002	0,7102	0,6083	0,5000
	4		1,0000	0,9998	0,9988	0,9953	0,9871	0,9712	0,9547	0,9444	0,9037	0,8471	0,7734
	5			1,0000	0,9999	0,9996	0,9987	0,9962	0,9931	0,9910	0,9812	0,9643	0,9375
	6				1,0000	1,0000	0,9999	0,9998	0,9995	0,9994	0,9984	0,9963	0,9922
	7						1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
8	**0**	0,9227	0,6634	0,4305	0,2725	0,1678	0,1001	0,0576	0,0390	0,0319	0,0168	0,0084	0,0039
	1	0,9973	0,9428	0,8131	0,6572	0,5033	0,3671	0,2553	0,1951	0,1691	0,1064	0,0632	0,0352
	2	0,9999	0,9942	0,9619	0,8948	0,7969	0,6785	0,5518	0,4682	0,4278	0,3154	0,2201	0,1445
	3	1,0000	0,9996	0,9950	0,9786	0,9437	0,8862	0,8059	0,7413	0,7064	0,5941	0,4770	0,3633
	4		1,0000	0,9996	0,9971	0,9896	0,9727	0,9420	0,9121	0,8939	0,8263	0,7396	0,6367
	5			1,0000	0,9998	0,9988	0,9958	0,9887	0,9803	0,9747	0,9502	0,9115	0,8555
	6				1,0000	0,9999	0,9996	0,9987	0,9974	0,9964	0,9915	0,9819	0,9648
	7					1,0000	1,0000	0,9999	0,9998	0,9998	0,9993	0,9983	0,9961
	8							1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Tabla 2 Probabilidades binomiales acumuladas *(continuación)*

		P										
n	**k**	**0,01**	**0,05**	**0,10**	**0,15**	**0,20**	**0,25**	**0,30**	**1/3**	**0,35**	**0,40**	**0,45**	**0,50**
9	**0**	0,9135	0,6302	0,3874	0,2316	0,1342	0,0751	0,0404	0,0260	0,0207	0,0101	0,0046	0,0020
	1	0,9965	0,9288	0,7748	0,5995	0,4362	0,3003	0,1960	0,1431	0,1211	0,0705	0,0385	0,0195
	2	0,9999	0,9916	0,9470	0,8591	0,7382	0,6007	0,4628	0,3772	0,3373	0,2318	0,1495	0,0898
	3	1,0000	0,9994	0,9917	0,9661	0,9144	0,8343	0,7297	0,6503	0,6089	0,4826	0,3614	0,2539
	4		1,0000	0,9991	0,9944	0,9804	0,9511	0,9012	0,8552	0,8283	0,7334	0,6214	0,5000
	5			0,9999	0,9994	0,9969	0,9900	0,9747	0,9576	0,9464	0,9006	0,8342	0,7461
	6			1,0000	1,0000	0,9997	0,9987	0,9957	0,9917	0,9888	0,9750	0,9502	0,9102
	7					1,0000	0,9999	0,9996	0,9990	0,9986	0,9962	0,9909	0,9805
	8						1,0000	1,0000	0,9999	0,9999	0,9997	0,9992	0,9980
	9								1,0000	1,0000	1,0000	1,0000	1,0000
10	**0**	0,9044	0,5987	0,3487	0,1969	0,1074	0,0563	0,0282	0,0173	0,0135	0,0060	0,0025	0,0010
	1	0,9958	0,9139	0,7361	0,5443	0,3758	0,2440	0,1493	0,1040	0,0860	0,0464	0,0233	0,0107
	2	1,0000	0,9885	0,9298	0,8202	0,6778	0,5256	0,3828	0,2991	0,2616	0,1673	0,0996	0,0547
	3		0,9990	0,9872	0,9500	0,8791	0,7759	0,6496	0,5593	0,5138	0,3823	0,2660	0,1719
	4		0,9999	0,9984	0,9901	0,9672	0,9219	0,8497	0,7869	0,7515	0,6331	0,5044	0,3770
	5		1,0000	0,9999	0,9986	0,9936	0,9803	0,9527	0,9234	0,9051	0,8338	0,7384	0,6230
	6			1,0000	0,9999	0,9991	0,9965	0,9894	0,9803	0,9740	0,9452	0,8980	0,8281
	7				1,0000	0,9999	0,9996	0,9984	0,9966	0,9952	0,9877	0,9726	0,9453
	8					1,0000	1,0000	0,9999	0,9996	0,9995	0,9983	0,9955	0,9893
	9							1,0000	1,0000	1,0000	0,9999	0,9997	0,9990
	10										1,0000	1,0000	1,0000

Tabla 3 Probabilidades de Poisson acumuladas

	λ													
k	**0,1**	**0,2**	**0,3**	**0,4**	**0,5**	**0,6**	**0,7**	**0,8**	**0,9**	**1,0**	**1,1**	**1,2**	**1,3**	**1,4**
0	0,905	0,819	0,741	0,670	0,607	0,549	0,497	0,449	0,407	0,368	0,333	0,301	0,273	0,247
1	0,995	0,982	0,963	0,938	0,910	0,878	0,844	0,809	0,772	0,736	0,699	0,663	0,627	0,592
2	1,000	0,999	0,996	0,992	0,986	0,977	0,966	0,953	0,937	0,920	0,900	0,879	0,857	0,833
3		1,000	1,000	0,999	0,998	0,997	0,994	0,991	0,987	0,981	0,974	0,966	0,957	0,946
4				1,000	1,000	1,000	0,999	0,999	0,998	0,996	0,995	0,992	0,989	0,986
5							1,000	1,000	1,000	0,999	0,999	0,998	0,998	0,997
6										1,000	1,000	1,000	1,000	0,999
7														1,000

k	**1,5**	**1,6**	**1,7**	**1,8**	**1,9**	**2,0**	**2,2**	**2,4**	**2,6**	**2,8**	**3,0**	**3,2**	**3,4**	**3,6**
0	0,223	0,202	0,183	0,165	0,150	0,135	0,111	0,091	0,074	0,061	0,050	0,041	0,033	0,027
1	0,558	0,525	0,493	0,463	0,434	0,406	0,355	0,308	0,267	0,231	0,199	0,171	0,147	0,126
2	0,809	0,783	0,757	0,731	0,704	0,677	0,623	0,570	0,518	0,469	0,423	0,380	0,340	0,303
3	0,934	0,921	0,907	0,891	0,875	0,857	0,819	0,779	0,736	0,692	0,647	0,603	0,558	0,515
4	0,981	0,976	0,970	0,964	0,956	0,947	0,928	0,904	0,877	0,848	0,815	0,781	0,744	0,706
5	0,996	0,994	0,992	0,990	0,987	0,983	0,975	0,964	0,951	0,935	0,916	0,895	0,871	0,844
6	0,999	0,999	0,998	0,997	0,997	0,995	0,993	0,988	0,983	0,976	0,966	0,955	0,942	0,927
7	1,000	1,000	1,000	0,999	0,999	0,999	0,998	0,997	0,995	0,992	0,988	0,983	0,977	0,969
8				1,000	1,000	1,000	1,000	0,999	0,999	0,998	0,996	0,994	0,992	0,988
9								1,000	1,000	0,999	0,999	0,998	0,997	0,996
10										1,000	1,000	1,000	0,999	0,999
11													1,000	1,000

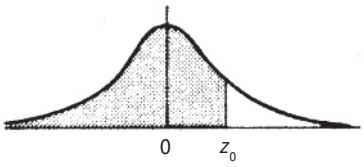
k	**3,8**	**4,0**	**4,2**	**4,4**	**4,6**	**4,8**	**5,0**	**5,2**	**5,4**	**5,6**	**5,8**	**6,0**
0	0,022	0,018	0,015	0,012	0,010	0,008	0,007	0,006	0,005	0,004	0,003	0,002
1	0,107	0,092	0,078	0,066	0,056	0,048	0,040	0,034	0,029	0,024	0,021	0,017
2	0,269	0,238	0,210	0,185	0,163	0,143	0,125	0,109	0,095	0,082	0,072	0,062
3	0,473	0,433	0,395	0,359	0,326	0,294	0,265	0,238	0,213	0,191	0,170	0,151
4	0,668	0,629	0,590	0,551	0,513	0,476	0,440	0,406	0,373	0,342	0,313	0,285
5	0,816	0,785	0,753	0,720	0,686	0,651	0,616	0,581	0,546	0,512	0,478	0,446
6	0,909	0,889	0,867	0,844	0,818	0,791	0,762	0,732	0,702	0,670	0,638	0,606
7	0,960	0,949	0,936	0,921	0,905	0,887	0,867	0,845	0,822	0,797	0,771	0,744
8	0,984	0,979	0,972	0,964	0,955	0,944	0,932	0,918	0,903	0,886	0,867	0,847
9	0,994	0,992	0,989	0,985	0,980	0,975	0,968	0,960	0,951	0,941	0,929	0,916
10	0,998	0,997	0,996	0,994	0,992	0,990	0,986	0,982	0,977	0,972	0,965	0,957
11	0,999	0,999	0,999	0,998	0,997	0,996	0,995	0,993	0,990	0,988	0,984	0,980
12	1,000	1,000	1,000	0,999	0,999	0,999	0,998	0,997	0,996	0,995	0,993	0,991
13				1,000	1,000	1,000	0,999	0,999	0,999	0,998	0,997	0,996
14							1,000	1,000	0,999	0,999	0,999	0,999
15									1,000	1,000	1,000	0,999
16												1,000

Tabla 4 Distribución normal estandarizada, $N(0,1)$

Valores de la función de distribución,

$$\Phi(z_0) = p(z \leq z_0)$$

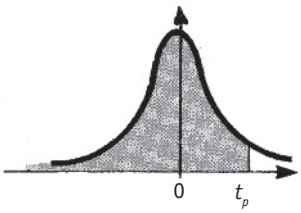
La tabla da el área sombreada en la figura



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7703	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,90147
1,3	0,90320	0,90490	0,90658	0,90824	0,90988	0,91149	0,91309	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,92220	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,94520	0,94630	0,94738	0,94845	0,94950	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,96080	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,97320	0,97381	0,97441	0,97500	0,97558	0,97615	0,97670
2,0	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,98030	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,98300	0,98341	0,98382	0,98422	0,98461	0,98500	0,98537	0,98574
2,2	0,98610	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,98840	0,98870	0,98899
2,3	0,98928	0,98956	0,98983	$0,9^20097$	$0,9^20358$	$0,9^20613$	$0,9^20863$	$0,9^21106$	$0,9^21344$	$0,9^21576$
2,4	$0,9^21802$	$0,9^22024$	$0,9^22240$	$0,9^22451$	$0,9^22656$	$0,9^22857$	$0,9^23053$	$0,9^23244$	$0,9^23431$	$0,9^23613$
2,5	$0,9^23790$	$0,9^23963$	$0,9^24132$	$0,9^24297$	$0,9^24457$	$0,9^24614$	$0,9^24766$	$0,9^24915$	$0,9^25060$	$0,9^25201$
2,6	$0,9^25339$	$0,9^25473$	$0,9^25604$	$0,9^25731$	$0,9^25855$	$0,9^25975$	$0,9^26093$	$0,9^26207$	$0,9^26319$	$0,9^26427$
2,7	$0,9^26533$	$0,9^26636$	$0,9^26736$	$0,9^26833$	$0,9^26928$	$0,9^27020$	$0,9^27110$	$0,9^27197$	$0,9^27282$	$0,9^27365$
2,8	$0,9^27445$	$0,9^27523$	$0,9^27599$	$0,9^27673$	$0,9^27744$	$0,9^27814$	$0,9^27882$	$0,9^27948$	$0,9^28012$	$0,9^28074$
2,9	$0,9^28134$	$0,9^28193$	$0,9^28250$	$0,9^28305$	$0,9^28359$	$0,9^28411$	$0,9^28462$	$0,9^28511$	$0,9^28559$	$0,9^28605$
3,0	$0,9^28650$	$0,9^28694$	$0,9^28736$	$0,9^28777$	$0,9^28817$	$0,9^28856$	$0,9^28893$	$0,9^28930$	$0,9^28965$	$0,9^28999$
3,1	$0,9^30324$	$0,9^30646$	$0,9^30957$	$0,9^31260$	$0,9^31553$	$0,9^31836$	$0,9^32112$	$0,9^32378$	$0,9^32636$	$0,9^32886$
3,2	$0,9^33129$	$0,9^33363$	$0,9^33590$	$0,9^33810$	$0,9^34024$	$0,9^34230$	$0,9^34429$	$0,9^34623$	$0,9^34810$	$0,9^34991$
3,3	$0,9^35166$	$0,9^35335$	$0,9^35499$	$0,9^35658$	$0,9^35811$	$0,9^35959$	$0,9^36103$	$0,9^36242$	$0,9^36376$	$0,9^36505$
3,4	$0,9^36631$	$0,9^36752$	$0,9^36869$	$0,9^36982$	$0,9^37091$	$0,9^37197$	$0,9^37299$	$0,9^37398$	$0,9^37493$	$0,9^37585$
3,5	$0,9^37674$	$0,9^37759$	$0,9^37842$	$0,9^37922$	$0,9^37999$	$0,9^38074$	$0,9^38146$	$0,9^38215$	$0,9^38282$	$0,9^38347$
3,6	$0,9^38409$	$0,9^38469$	$0,9^38527$	$0,9^38583$	$0,9^38637$	$0,9^38689$	$0,9^38739$	$0,9^38787$	$0,9^38834$	$0,9^38879$
3,7	$0,9^38922$	$0,9^38964$	$0,9^40039$	$0,9^40426$	$0,9^40799$	$0,9^41158$	$0,9^41504$	$0,9^41838$	$0,9^42159$	$0,9^42468$
3,8	$0,9^42765$	$0,9^43052$	$0,9^43327$	$0,9^43593$	$0,9^43848$	$0,9^44094$	$0,9^44331$	$0,9^44558$	$0,9^44777$	$0,9^44988$
3,9	$0,9^45190$	$0,9^45385$	$0,9^45573$	$0,9^45753$	$0,9^45926$	$0,9^46092$	$0,9^46253$	$0,9^46406$	$0,9^46554$	$0,9^46696$
4,0	$0,9^46833$	$0,9^46964$	$0,9^47090$	$0,9^47211$	$0,9^47327$	$0,9^47439$	$0,9^47546$	$0,9^47649$	$0,9^47748$	$0,9^47843$

Tabla 5 Distribución *t* de Student

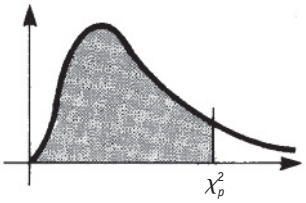
Valores de la función de distribución, $P(t \leq t_p) = p$; k = grados de libertad. La tabla da el área sombreada en la figura.



k	$t_{0,995}$	$t_{0,99}$	$t_{0,975}$	$t_{0,95}$	$t_{0,90}$	$t_{0,80}$	$t_{0,75}$	$t_{0,70}$	$t_{0,60}$	$t_{0,55}$
1	63,66	31,82	12,71	6,31	3,08	1,376	1,000	0,727	0,325	0,158
2	9,92	6,96	4,30	2,92	1,89	1,061	0,816	0,617	0,289	0,142
3	5,84	4,54	3,18	2,35	1,64	0,978	0,765	0,584	0,277	0,137
4	4,60	3,75	2,78	2,13	1,53	0,941	0,741	0,569	0,271	0,134
5	4,03	3,36	2,57	2,02	1,48	0,920	0,727	0,559	0,267	0,132
6	3,71	3,14	2,45	1,94	1,44	0,906	0,718	0,553	0,265	0,131
7	3,50	3,00	2,36	1,90	1,42	0,896	0,711	0,549	0,263	0,130
8	3,36	2,90	2,31	1,86	1,40	0,889	0,706	0,546	0,262	0,130
9	3,25	2,82	2,26	1,83	1,38	0,883	0,703	0,543	0,261	0,129
10	3,17	2,76	2,23	1,81	1,37	0,879	0,700	0,542	0,260	0,129
11	3,11	2,72	2,20	1,80	1,36	0,876	0,697	0,540	0,260	0,129
12	3,06	2,68	2,18	1,78	1,36	0,873	0,695	0,539	0,259	0,128
13	3,01	2,65	2,16	1,77	1,35	0,870	0,694	0,538	0,259	0,128
14	2,98	2,62	2,14	1,76	1,34	0,868	0,692	0,537	0,258	0,128
15	2,95	2,60	2,13	1,75	1,34	0,866	0,691	0,536	0,258	0,128
16	2,92	2,58	2,12	1,75	1,34	0,865	0,690	0,535	0,258	0,128
17	2,90	2,57	2,11	1,74	1,33	0,863	0,689	0,534	0,257	0,128
18	2,88	2,55	2,10	1,73	1,33	0,862	0,688	0,534	0,257	0,127
19	2,86	2,54	2,09	1,73	1,33	0,861	0,688	0,533	0,257	0,127
20	2,84	2,53	2,09	1,72	1,32	0,860	0,687	0,533	0,257	0,127
21	2,83	2,52	2,08	1,72	1,32	0,859	0,686	0,532	0,257	0,127
22	2,82	2,51	2,07	1,72	1,32	0,858	0,686	0,532	0,256	0,127
23	2,81	2,50	2,07	1,71	1,32	0,858	0,685	0,532	0,256	0,127
24	2,80	2,49	2,06	1,71	1,32	0,857	0,685	0,531	0,256	0,127
25	2,79	2,48	2,06	1,71	1,32	0,856	0,684	0,531	0,256	0,127
26	2,78	2,48	2,06	1,71	1,32	0,856	0,684	0,531	0,256	0,127
27	2,77	2,47	2,05	1,70	1,31	0,855	0,684	0,531	0,256	0,127
28	2,76	2,47	2,05	1,70	1,31	0,855	0,683	0,530	0,256	0,127
29	2,76	2,46	2,04	1,70	1,31	0,854	0,683	0,530	0,256	0,127
30	2,75	2,46	2,04	1,70	1,31	0,854	0,683	0,530	0,256	0,127
40	2,70	2,42	2,02	1,68	1,30	0,851	0,681	0,529	0,255	0,126
60	2,66	2,39	2,00	1,67	1,30	0,848	0,679	0,527	0,254	0,126
120	2,62	2,36	1,98	1,66	1,29	0,845	0,677	0,526	0,254	0,126
∞	2,58	2,33	1,96	1,645	1,28	0,842	0,674	0,524	0,253	0,126

Tabla 6 Distribución *chi*-cuadrado de Pearson

Valores de la función de distribución, $P(\chi^2 \leq \chi^2_p) = p$; k = grados de libertad. (Área sombreada en la figura). Para $k > 100$ utilizar que $\sqrt{2\chi^2_k} \sim N\sqrt{2k-1}; 1)$.



k	$\chi^2_{0,995}$	$\chi^2_{0,99}$	$\chi^2_{0,975}$	$\chi^2_{0,95}$	$\chi^2_{0,90}$	$\chi^2_{0,75}$	$\chi^2_{0,50}$	$\chi^2_{0,25}$	$\chi^2_{0,10}$	$\chi^2_{0,05}$	$\chi^2_{0,025}$	$\chi^2_{0,01}$	$\chi^2_{0,005}$
1	7,88	6,63	5,02	3,84	2,71	1,32	0,455	0,102	0,0158	0,0039	0,0010	0,0002	0,0000
2	10,6	9,21	7,38	5,99	4,61	2,77	1,39	0,575	0,211	0,103	0,0506	0,0201	0,0100
3	12,8	11,3	9,35	7,81	6,25	4,11	2,37	1,21	0,584	0,352	0,216	0,115	0,072
4	14,9	13,3	11,1	9,49	7,78	5,39	3,36	1,92	1,06	0,711	0,484	0,297	0,207
5	16,7	15,1	12,8	11,1	9,24	6,63	4,35	2,67	1,61	1,15	0,831	0,554	0,412
6	18,5	16,8	14,4	12,6	10,6	7,84	5,35	3,45	2,20	1,64	1,24	0,872	0,676
7	20,3	18,5	16,0	14,1	12,0	9,04	6,35	4,25	2,83	2,17	1,69	1,24	0,989
8	22,0	20,1	17,5	15,5	13,4	10,2	7,34	5,07	3,49	2,73	2,18	1,65	1,34
9	23,6	21,7	19,0	16,9	14,7	11,4	8,34	5,90	4,17	3,33	2,70	2,09	1,73
10	25,2	23,2	20,5	18,3	16,0	12,5	9,34	6,74	4,87	3,94	3,25	2,56	2,16
11	26,8	24,7	21,9	19,7	17,3	13,7	10,3	7,58	5,58	4,57	3,82	3,05	2,60
12	28,3	26,2	23,3	21,0	18,5	14,8	11,3	8,44	6,30	5,23	4,40	3,57	3,07
13	29,8	27,7	24,7	22,4	19,8	16,0	12,3	9,30	7,04	5,89	5,01	4,11	3,57
14	31,3	29,1	26,1	23,7	21,1	17,1	13,3	10,2	7,79	6,57	5,63	4,66	4,07
15	32,8	30,6	27,5	25,0	22,3	18,2	14,3	11,0	8,55	7,26	6,26	5,23	4,60
16	34,3	32,0	28,8	26,3	23,5	19,4	15,3	11,9	9,31	7,96	6,91	5,81	5,14
17	35,7	33,4	30,2	27,6	24,8	20,5	16,3	12,8	10,1	8,67	7,56	6,41	5,70
18	37,2	34,8	31,5	28,9	26,0	21,6	17,3	13,7	10,9	9,39	8,23	7,01	6,26
19	38,6	36,2	32,9	30,1	27,2	22,7	18,3	14,6	11,7	10,1	8,91	7,63	6,84
20	40,0	37,6	34,2	31,4	28,4	23,8	19,3	15,5	12,4	10,9	9,59	8,26	7,43
21	41,4	38,9	35,5	32,7	29,6	24,9	20,3	16,3	13,2	11,6	10,3	8,90	8,03
22	42,8	40,3	36,8	33,9	30,8	26,0	21,3	17,2	14,0	12,3	11,0	9,54	8,64
23	44,2	41,6	38,1	35,2	32,0	27,1	22,3	18,1	14,8	13,1	11,7	10,2	9,26
24	45,6	43,0	39,4	36,4	33,2	28,2	23,3	19,0	15,7	13,8	12,4	10,9	9,89
25	46,9	44,3	40,6	37,7	34,4	29,3	24,3	19,9	16,5	14,6	13,1	11,5	10,5
26	48,3	45,6	41,9	38,9	35,6	30,4	25,3	20,8	17,3	15,4	13,8	12,2	11,2
27	49,6	47,0	43,2	40,1	36,7	31,5	26,3	21,7	18,1	16,2	14,6	12,9	11,8
28	51,0	48,3	44,5	41,3	37,9	32,6	27,3	22,7	18,9	16,9	15,3	13,6	12,5
29	52,3	49,6	45,7	42,6	39,1	33,7	28,3	33,6	19,8	17,7	16,0	14,3	13,1
30	53,7	50,9	47,0	43,8	40,3	34,8	29,3	24,5	20,6	18,5	16,8	15,0	13,8
40	66,8	63,7	59,3	55,8	51,8	45,6	39,3	33,7	29,1	26,5	24,4	22,2	20,7
50	79,5	76,2	71,4	67,5	63,2	56,3	49,3	42,9	37,7	34,8	32,4	29,7	28,0
60	92,0	88,4	83,3	79,1	74,4	67,0	59,3	52,3	46,5	43,2	40,5	37,5	35,5
70	104,2	100,4	95,0	90,5	85,5	77,6	69,3	61,7	55,3	51,7	48,8	45,4	43,3
80	116,3	112,3	106,6	101,9	96,6	88,1	79,3	71,1	64,3	60,4	57,2	53,5	51,2
90	128,3	124,1	118,1	113,1	107,6	98,6	89,3	80,6	73,3	69,1	65,6	61,8	59,2
100	140,2	135,8	129,6	124,3	118,5	109,1	99,3	90,1	82,4	77,9	74,2	70,1	67,3

Tabla 7a Distribución *F*

n_1 grados de libertad del numerador, n_2 grados de libertad del denominador $\alpha = 0{,}05$

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	243,9	245,9	248,0	249,1	250,1	251,1	252,2	253,3	254,3
2	18,51	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,43	19,45	19,45	19,46	19,47	19,48	19,48	19,49	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13
15	5,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,13	2,05	2,01	1,96	1,91	1,86	1,81	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79	1,73
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77	1,71
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,07	1,99	1,95	1,90	1,85	1,80	1,75	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,13	2,06	1,97	1,95	1,88	1,84	1,79	1,73	1,67
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,04	1,96	1,91	1,87	1,82	1,77	1,71	1,65
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,10	2,03	1,94	1,90	1,85	1,81	1,75	1,70	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58	1,51
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35	1,25
∞	3,94	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22	1,00

Tabla 7b Distribución *F*

$\alpha = 0{,}01$

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4.052	4.999,5	5.403	5.625	5.764	5.859	5.928	5.982	6.022	6.056	6.101	6.157	6.209	6.235	6.261	6.287	6.313	6.339	6.366
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39	99,40	99,42	99,43	99,45	99,46	99,47	99,47	99,48	99,49	99,50
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35	27,23	27,05	26,87	26,69	26,60	26,50	26,41	26,32	26,22	26,13
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	14,37	14,20	14,02	13,93	13,84	13,75	13,65	13,56	13,46
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,89	9,72	9,55	9,47	9,38	9,29	9,20	9,11	9,02
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,72	7,56	7,40	7,31	7,23	7,14	7,06	6,97	6,88
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,47	6,31	6,16	6,07	5,99	5,91	5,82	5,74	5,65
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,67	5,52	5,36	5,28	5,20	5,12	5,03	4,95	4,86
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,11	4,96	4,81	4,73	4,65	4,57	4,48	4,40	4,31
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,71	4,56	4,41	4,33	4,25	4,17	4,08	4,00	3,91
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,40	4,25	4,10	4,02	3,94	3,86	3,78	3,69	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,16	4,01	3,86	3,78	3,70	3,62	3,54	3,45	3,36
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	3,96	3,82	3,66	3,59	3,51	3,43	3,34	3,25	3,17
14	9,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,80	3,66	3,51	3,43	3,35	3,27	3,18	3,09	3,00
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,67	3,52	3,37	3,29	3,21	3,13	3,05	2,96	2,87
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,55	3,41	3,26	3,18	3,10	3,02	2,93	2,84	2,75
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,46	3,31	3,16	3,08	3,00	2,92	2,83	2,75	2,65
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,37	3,23	3,08	3,00	2,92	2,84	2,75	2,66	2,57
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,30	3,15	3,00	2,92	2,84	2,76	2,67	2,58	2,49
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,23	3,09	2,94	2,86	2,78	2,69	2,61	2,52	2,42
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,17	3,03	2,88	2,80	2,72	2,64	2,55	2,46	2,36
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,12	2,98	2,83	2,75	2,67	2,58	2,50	2,40	2,31
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,07	2,93	2,78	2,70	2,62	2,54	2,45	2,35	2,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,03	2,89	2,74	2,66	2,58	2,49	2,40	2,31	2,21
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13	2,99	2,85	2,70	2,62	2,54	2,45	2,36	2,27	2,17
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09	2,96	2,81	2,66	2,58	2,50	2,42	2,33	2,23	2,13
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06	2,93	2,78	2,63	2,55	2,47	2,38	2,29	2,20	2,10
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03	2,90	2,75	2,60	2,52	2,44	2,35	2,26	2,17	2,06
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00	2,87	2,73	2,57	2,49	2,41	2,33	2,23	2,14	2,03
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,84	2,80	2,55	2,47	2,39	2,30	2,21	2,11	2,01
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,66	2,52	2,37	2,29	2,20	2,11	2,02	1,92	1,80
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,50	2,35	2,20	2,12	2,03	1,94	1,84	1,73	1,60
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47	2,34	2,19	2,03	1,95	1,86	1,76	1,66	1,53	1,38
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	2,18	2,04	1,88	1,79	1,70	1,59	1,47	1,32	1,00

Tabla 8 Contraste de Kolmogorov-Smirnov

Valores críticos de $D = |F_n(x) - F(x)|$ donde $F_n(x)$ es la distribución muestral de tamaño n y $F(x)$ la distribución teórica.

Tamaño muestral n	Nivel de significación				
	0,20	**0,15**	**0,10**	**0,05**	**0,01**
1	0,900	0,925	0,950	0,975	0,995
2	0,684	0,726	0,776	0,842	0,929
3	0,565	0,597	0,642	0,708	0,828
4	0,494	0,525	0,564	0,624	0,733
5	0,446	0,474	0,510	0,565	0,669
6	0,410	0,436	0,470	0,521	0,618
7	0,381	0,405	0,438	0,486	0,577
8	0,358	0,381	0,411	0,457	0,543
9	0,339	0,360	0,388	0,432	0,514
10	0,322	0,342	0,368	0,410	0,490
11	0,307	0,326	0,352	0,391	0,468
12	0,295	0,313	0,338	0,375	0,450
13	0,284	0,302	0,325	0,361	0,433
14	0,274	0,292	0,314	0,349	0,418
15	0,266	0,283	0,304	0,338	0,404
16	0,258	0,274	0,295	0,328	0,392
17	0,250	0,266	0,286	0,318	0,381
18	0,244	0,259	0,278	0,309	0,371
19	0,237	0,252	0,272	0,301	0,363
20	0,231	0,246	0,264	0,294	0,356
25	0,21	0,22	0,24	0,27	0,32
30	0,19	0,20	0,22	0,24	0,29
35	0,18	0,19	0,21	0,23	0,27
>35	$\frac{1,07}{\sqrt{n}}$	$\frac{1,14}{\sqrt{n}}$	$\frac{1,22}{\sqrt{n}}$	$\frac{1,36}{\sqrt{n}}$	$\frac{1,63}{\sqrt{n}}$

n es el tamaño de la muestra.

Tabla 9 Contraste Kolmogorov-Smirnov (Lilliefors)

Tablas de $D_n = |F_n(x) - F(x)|$ para contrastar la hipótesis de normalidad cuando la media y la varianza poblacionales son estimadas por sus valores muestrales.

Tamaño muestral *n*	Nivel de significación				
	0,20	**0,15**	**0,10**	**0,05**	**0,01**
4	0,300	0,319	0,352	0,381	0,417
5	0,285	0,299	0,315	0,337	0,405
6	0,265	0,277	0,294	0,319	0,364
7	0,247	0,258	0,276	0,300	0,348
8	0,233	0,244	0,261	0,285	0,331
9	0,223	0,233	0,249	0,271	0,311
10	0,215	0,224	0,239	0,258	0,294
11	0,206	0,217	0,230	0,249	0,284
12	0,199	0,212	0,223	0,242	0,275
13	0,190	0,202	0,214	0,234	0,268
14	0,183	0,194	0,207	0,227	0,261
15	0,177	0,187	0,201	0,220	0,257
16	0,173	0,182	0,195	0,213	0,250
17	0,169	0,177	0,189	0,206	0,245
18	0,166	0,173	0,184	0,200	0,239
19	0,163	0,169	0,179	0,195	0,235
20	0,160	0,166	0,174	0,190	0,231
25	0,149	0,153	0,165	0,180	0,203
30	0,131	0,136	0,144	0,161	0,187
>30	$\frac{0,736}{\sqrt{n}}$	$\frac{0,768}{\sqrt{n}}$	$\frac{0,805}{\sqrt{n}}$	$\frac{0,886}{\sqrt{n}}$	$\frac{1,031}{\sqrt{n}}$

Tabla 10 Coeficientes del contraste de Shapiro-Wilk

Coeficientes a_{in} para el contraste W de Shapiro y Wilk, n es el tamaño muestral.

i \ n	2	3	4	5	6	7	8	9	10
1	0,7071	0,7071	0,6872	0,6646	0,6431	0,6233	0,6052	0,5888	0,5739
2	—	0,0000	0,1677	0,2413	0,2806	0,3031	0,3164	0,3244	0,3291
3	—	—	—	0,0000	0,0875	0,1401	0,1743	0,1976	0,2141
4	—	—	—	—	—	0,0000	0,0561	0,0947	0,1224
5	—	—	—	—	—	—	—	0,0000	0,0399

i \ n	11	12	13	14	15	16	17	18	19	20
1	0,5601	0,5475	0,5359	0,5251	0,5150	0,5056	0,4968	0,4886	0,4808	0,4734
2	0,3315	0,3325	0,3325	0,3318	0,3306	0,3290	0,3273	0,3253	0,3232	0,3211
3	0,2260	0,2347	0,2412	0,2495	0,2495	0,2521	0,2540	0,2553	0,2561	0,2565
4	0,1429	0,1586	0,1707	0,1802	0,1878	0,1988	0,1988	0,2027	0,2059	0,2085
5	0,0695	0,0922	0,1099	0,1240	0,1353	0,1447	0,1524	0,1587	0,1641	0,1686
6	0,0000	0,0303	0,0539	0,0727	0,0880	0,1005	0,1109	0,1197	0,1271	0,1334
7	—	—	0,0000	0,0240	0,0433	0,0593	0,0725	0,0837	0,0932	0,1013
8	—	—	—	—	0,0000	0,0196	0,0359	0,0496	0,0612	0,0711
9	—	—	—	—	—	—	0,0000	0,0163	0,0303	0,0422
10	—	—	—	—	—	—	—	—	0,0000	0,0140

i \ n	21	22	23	24	25	26	27	28	29	30
1	0,4643	0,4590	0,4542	0,4493	0,4450	0,4407	0,4366	0,4328	0,4291	0,4254
2	0,3185	0,3156	0,3126	0,3098	0,3069	0,3043	0,3018	0,2992	0,2968	0,2944
3	0,2578	0,2571	0,2563	0,2554	0,2543	0,2533	0,2522	0,2510	0,2499	0,2487
4	0,2119	0,2131	0,2139	0,2145	0,2148	0,2151	0,2152	0,2151	0,2150	0,2148
5	0,1736	0,1764	0,1787	0,1807	0,1822	0,1836	0,1848	0,1857	0,1864	0,1870
6	0,1399	0,1443	0,1480	0,1512	0,1539	0,1563	0,1584	0,1601	0,1616	0,1630
7	0,1092	0,1150	0,1201	0,1245	0,1283	0,1316	0,1346	0,1372	0,1395	0,1415
8	0,0804	0,0878	0,0941	0,0997	0,1046	0,1089	0,1128	0,1162	0,1192	0,1219
9	0,0530	0,0618	0,0696	0,0764	0,0823	0,0876	0,0923	0,0965	0,1002	0,1036
10	0,0263	0,0368	0,0459	0,0539	0,0610	0,0672	0,0728	0,0778	0,0822	0,0862
11	0,0000	0,0122	0,0228	0,0321	0,0403	0,0476	0,0540	0,0598	0,0650	0,0697
12	—	—	0,0000	0,0107	0,0200	0,0284	0,0358	0,0424	0,0483	0,0537
13	—	—	—	—	0,0000	0,0094	0,0178	0,0253	0,0320	0,0381
14	—	—	—	—	—	—	0,0000	0,0084	0,0159	0,0227
15	—	—	—	—	—	—	—	—	0,0000	0,0076

Tabla 10 Coeficientes del contraste de Shapiro-Wilk *(continuación)*

i \ n	31	32	33	34	35	36	37	38	39	40
1	0,4220	0,4188	0,4156	0,4127	0,4096	0,4068	0,4040	0,4015	0,3989	0,3964
2	0,2921	0,2898	0,2876	0,2854	0,2834	0,2813	0,2794	0,2774	0,2755	0,2737
3	0,2475	0,2463	0,2451	0,2439	0,2427	0,2415	0,2403	0,2391	0,2380	0,2368
4	0,2145	0,2141	0,2137	0,2132	0,2127	0,2121	0,2116	0,2110	0,2104	0,2098
5	0,1874	0,1878	0,1880	0,1882	0,1883	0,1883	0,1883	0,1881	0,1880	0,1878
6	0,1641	0,1651	0,1660	0,1667	0,1673	0,1678	0,1683	0,1686	0,1689	0,1691
7	0,1433	0,1449	0,1463	0,1475	0,1487	0,1496	0,1505	0,1513	0,1520	0,1526
8	0,1243	0,1265	0,1284	0,1301	0,1317	0,1331	0,1344	0,1356	0,1366	0,1376
9	0,1066	0,1093	0,1118	0,1140	0,1160	0,1179	0,1196	0,1211	0,1225	0,1237
10	0,0899	0,0931	0,0961	0,0988	0,1013	0,1036	0,1056	0,1075	0,1092	0,1108
11	0,0739	0,0777	0,0812	0,0844	0,0873	0,0900	0,0924	0,0947	0,0967	0,0986
12	0,0585	0,0629	0,0669	0,0706	0,0739	0,0770	0,0798	0,0824	0,0848	0,0870
13	0,0435	0,0485	0,0530	0,0572	0,0610	0,0645	0,0677	0,0706	0,0733	0,0759
14	0,0289	0,0344	0,0395	0,0441	0,0484	0,0523	0,0559	0,0592	0,0622	0,0651
15	0,0144	0,0206	0,0262	0,0314	0,0361	0,0404	0,0444	0,0481	0,0515	0,0546
16	0,0000	0,0068	0,0187	0,0187	0,0239	0,0287	0,0331	0,0372	0,0409	0,0444
17	—	—	0,0000	0,0062	0,0119	0,0172	0,0220	0,0264	0,0305	0,0343
18	—	—	—	—	0,0000	0,0057	0,0110	0,0158	0,0203	0,0244
19	—	—	—	—	—	—	0,0000	0,0053	0,0101	0,0146
20	—	—	—	—	—	—	—	—	0,0000	0,0049

i \ n	41	42	43	44	45	46	47	48	49	50
1	0,3940	0,3917	0,3894	0,3872	0,3850	0,3830	0,3808	0,3789	0,3770	0,3751
2	0,2719	0,2701	0,2684	0,2667	0,2651	0,2635	0,2620	0,2604	0,2589	0,2574
3	0,2357	0,2345	0,2334	0,2323	0,2313	0,2302	0,2291	0,2281	0,2271	0,2260
4	0,2091	0,2085	0,2078	0,2072	0,2065	0,2058	0,2052	0,2045	0,2038	0,2032
5	0,1876	0,1874	0,1871	0,1868	0,1865	0,1862	0,1859	0,1855	0,1851	0,1847
6	0,1693	0,1694	0,1695	0,1695	0,1695	0,1695	0,1695	0,1693	0,1692	0,1691
7	0,1531	0,1535	0,1539	0,1542	0,1545	0,1548	0,1550	0,1551	0,1553	0,1554
8	0,1384	0,1392	0,1398	0,1405	0,1410	0,1415	0,1420	0,1423	0,1427	0,1430
9	0,1249	0,1259	0,1269	0,1278	0,1286	0,1293	0,1300	0,1306	0,1312	0,1317
10	0,1123	0,1136	0,1149	0,1160	0,1170	0,1180	0,1189	0,1197	0,1205	0,1212
11	0,1004	0,1020	0,1035	0,1049	0,1062	0,1073	0,1085	0,1095	0,1105	0,1113
12	0,0891	0,0909	0,0927	0,0943	0,0959	0,0972	0,0986	0,0998	0,1010	0,1020
13	0,0782	0,0804	0,0824	0,0842	0,0860	0,0876	0,0892	0,0906	0,0919	0,0932
14	0,0677	0,0701	0,0724	0,0745	0,0765	0,0783	0,0801	0,0817	0,0832	0,0846
15	0,0575	0,0602	0,0628	0,0651	0,0673	0,0694	0,0713	0,0731	0,0748	0,0764
16	0,0476	0,0506	0,0534	0,0560	0,0584	0,0607	0,0628	0,0648	0,0667	0,0685
17	0,0379	0,0411	0,0442	0,0471	0,0497	0,0522	0,0546	0,0568	0,0588	0,0608
18	0,0283	0,0318	0,0352	0,0383	0,0412	0,0439	0,0465	0,0489	0,0511	0,0532
19	0,0188	0,0227	0,0263	0,0296	0,0328	0,0357	0,0385	0,0411	0,0436	0,0459
20	0,0094	0,0136	0,0175	0,0211	0,0245	0,0277	0,0307	0,0335	0,0361	0,0386
21	0,0000	0,0045	0,0087	0,0126	0,0163	0,0197	0,0229	0,0259	0,0288	0,0314
22	—	—	0,0000	0,0042	0,0081	0,0118	0,0153	0,0185	0,0215	0,0244
23	—	—	—	—	0,0000	0,0039	0,0076	0,0111	0,0143	0,0174
24	—	—	—	—	—	—	0,0000	0,0037	0,0071	0,0104
25	—	—	—	—	—	—	—	—	0,0000	0,0035

Tabla 11 Percentiles del estadístico *W* de Shapiro y Wilk

La tabla proporciona los percentiles del estadístico *W* de Shapiro y Wilk en función del tamaño muestral, *n*.

	Nivel de significación								
n	**0,01**	**0,02**	**0,05**	**0,10**	**0,50**	**0,90**	**0,95**	**0,98**	**0,99**
3	0,753	0,756	0,767	0,789	0,959	0,998	0,999	1,000	1,000
4	0,687	0,707	0,748	0,792	0,935	0,987	0,992	0,996	0,997
5	0,686	0,715	0,762	0,806	0,927	0,979	0,986	0,991	0,993
6	0,713	0,743	0,788	0,826	0,927	0,974	0,981	0,986	0,989
7	0,730	0,760	0,803	0,838	0,928	0,972	0,979	0,985	0,988
8	0,749	0,778	0,818	0,851	0,932	0,972	0,978	0,984	0,987
9	0,764	0,791	0,829	0,859	0,935	0,972	0,978	0,984	0,986
10	0,781	0,806	0,842	0,869	0,938	0,972	0,978	0,983	0,986
11	0,792	0,817	0,850	0,876	0,940	0,973	0,979	0,984	0,986
12	0,805	0,828	0,859	0,883	0,943	0,973	0,979	0,984	0,986
13	0,814	0,837	0,866	0,889	0,945	0,974	0,979	0,984	0,986
14	0,825	0,846	0,874	0,895	0,947	0,975	0,980	0,984	0,986
15	0,835	0,855	0,881	0,901	0,950	0,975	0,980	0,984	0,987
16	0,844	0,863	0,887	0,906	0,952	0,976	0,981	0,985	0,987
17	0,851	0,869	0,892	0,910	0,954	0,977	0,981	0,985	0,987
18	0,858	0,874	0,897	0,914	0,956	0,978	0,982	0,986	0,988
19	0,863	0,879	0,901	0,917	0,957	0,978	0,982	0,986	0,988
20	0,868	0,884	0,905	0,920	0,959	0,979	0,983	0,986	0,988
21	0,873	0,888	0,908	0,923	0,960	0,980	0,983	0,987	0,989
22	0,878	0,892	0,911	0,926	0,961	0,980	0,984	0,987	0,989
23	0,881	0,895	0,914	0,928	0,962	0,981	0,984	0,987	0,989
24	0,884	0,898	0,916	0,930	0,963	0,981	0,984	0,987	0,989
25	0,888	0,901	0,918	0,931	0,964	0,981	0,985	0,988	0,989

Tabla 11 Percentiles del estadístico *W* de Shapiro y Wilk *(cont.)*

	Nivel de significación								
n	**0,01**	**0,02**	**0,05**	**0,10**	**0,50**	**0,90**	**0,95**	**0,98**	**0,99**
26	0,891	0,904	0,920	0,933	0,965	0,982	0,985	0,988	0,989
27	0,894	0,906	0,923	0,935	0,965	0,982	0,985	0,988	0,990
28	0,896	0,908	0,924	0,936	0,966	0,982	0,985	0,988	0,990
29	0,898	0,910	0,926	0,937	0,966	0,982	0,985	0,988	0,990
30	0,900	0,912	0,927	0,939	0,967	0,983	0,985	0,988	0,990
31	0,902	0,914	0,929	0,940	0,967	0,983	0,986	0,988	0,990
32	0,904	0,915	0,930	0,941	0,968	0,983	0,986	0,988	0,990
33	0,906	0,917	0,931	0,942	0,968	0,983	0,986	0,989	0,990
34	0,908	0,919	0,933	0,943	0,969	0,983	0,986	0,989	0,990
35	0,910	0,920	0,934	0,944	0,969	0,984	0,986	0,989	0,990
36	0,912	0,922	0,935	0,945	0,970	0,984	0,986	0,989	0,990
37	0,914	0,924	0,936	0,946	0,970	0,984	0,987	0,989	0,990
38	0,916	0,925	0,938	0,947	0,971	0,984	0,987	0,989	0,990
39	0,917	0,927	0,939	0,948	0,971	0,984	0,987	0,989	0,991
40	0,919	0,928	0,940	0,949	0,972	0,985	0,987	0,989	0,991
41	0,920	0,929	0,941	0,950	0,972	0,985	0,987	0,989	0,991
42	0,922	0,930	0,942	0,951	0,972	0,985	0,987	0,989	0,991
43	0,923	0,932	0,943	0,951	0,973	0,985	0,987	0,990	0,991
44	0,924	0,933	0,944	0,952	0,973	0,985	0,987	0,990	0,991
45	0,926	0,934	0,945	0,953	0,973	0,985	0,988	0,990	0,991
46	0,927	0,935	0,945	0,953	0,974	0,985	0,988	0,990	0,991
47	0,928	0,936	0,946	0,954	0,974	0,985	0,988	0,990	0,991
48	0,929	0,937	0,947	0,954	0,974	0,985	0,988	0,990	0,991
49	0,929	0,937	0,947	0,955	0,974	0,985	0,988	0,990	0,991
50	0,930	0,938	0,947	0,955	0,974	0,985	0,988	0,990	0,991

Tabla 12 Test de rachas

Percentiles de la distribución del número de rachas (r) en la hipótesis de independencia. El número de signos más —igual al de signos menos— es k. Para $k > 50$ utilizar la aproximación siguiente: el número de rachas es normal con parámetros:

$$\mu = k + 1$$

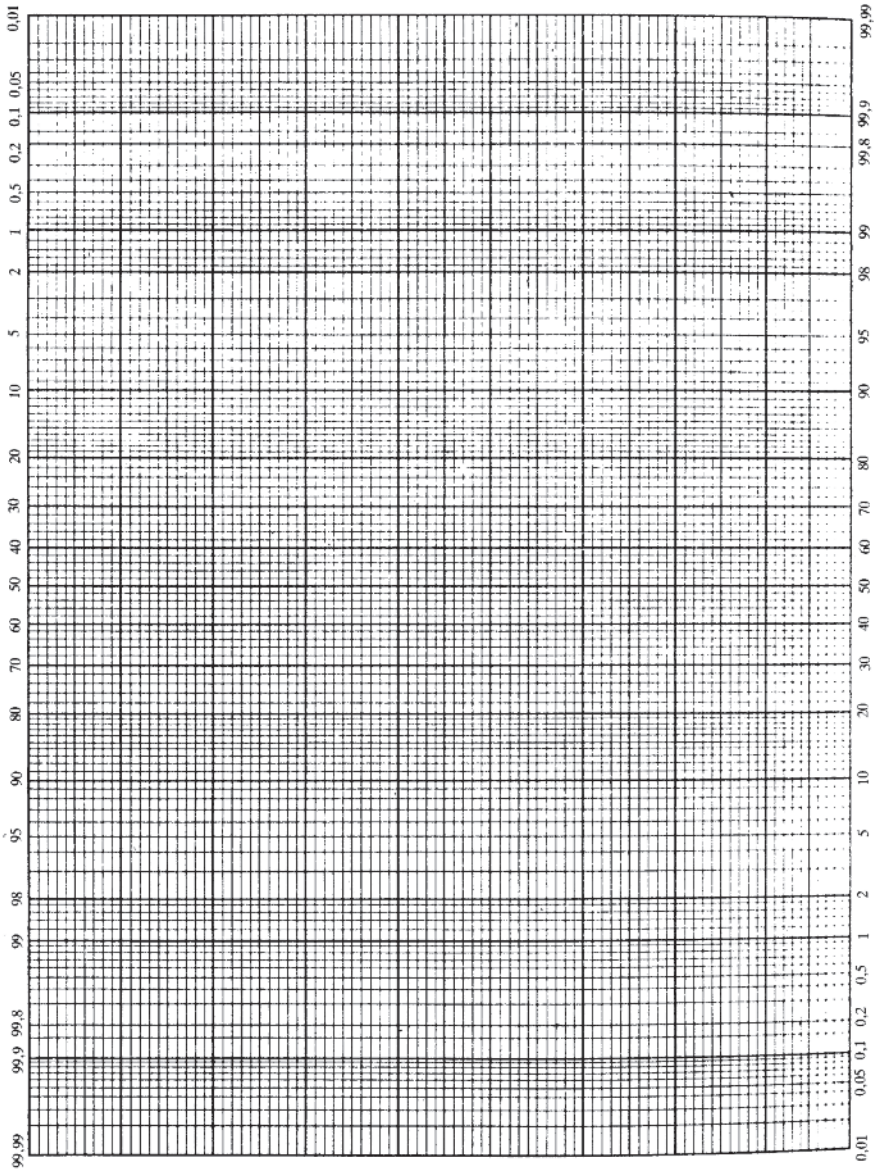
$$\sigma^2 = \frac{k(k-1)}{2k-1}$$

k	$r_{0,005}$	$r_{0,01}$	$r_{0,025}$	$r_{0,05}$	$r_{0,95}$	$r_{0,975}$	$r_{0,990}$	$r_{0,995}$
2	—	—	—	—	4	4	4	4
3	—	—	—	—	6	6	6	6
4	—	—	—	2	7	8	8	8
5	—	2	2	3	8	9	9	10
6	2	2	3	3	10	10	11	11
7	3	3	3	4	11	12	12	12
8	3	4	4	5	12	13	13	14
9	4	4	5	6	13	14	15	15
10	5	5	6	6	15	15	16	16
11	5	6	7	7	16	16	17	18
12	6	7	7	8	17	18	18	19
13	7	7	8	9	18	19	20	20
14	7	8	9	10	19	20	21	22
15	8	9	10	11	20	21	22	23
16	9	10	11	11	22	22	23	24
17	10	10	11	12	23	24	25	25
18	10	11	12	13	24	25	26	26
19	11	12	13	14	25	26	27	28
20	12	13	14	15	26	27	28	29
21	13	14	15	16	27	28	29	30
22	14	14	16	17	28	29	31	31
23	14	15	16	17	30	31	32	33
24	15	16	17	18	31	32	33	34

Tabla 12 Test de rachas *(continuación)*

k	$r_{0,005}$	$r_{0,01}$	$r_{0,025}$	$r_{0,05}$	$r_{0,95}$	$r_{0,975}$	$r_{0,990}$	$r_{0,995}$
25	16	17	18	19	32	33	34	35
26	17	18	19	20	33	34	35	36
27	18	19	20	21	34	35	36	37
28	18	19	21	22	35	36	38	39
29	19	20	22	23	36	37	39	40
30	20	21	22	24	37	39	40	41
31	21	22	23	25	38	40	41	42
32	22	23	24	25	40	41	42	43
33	23	24	25	26	41	42	43	44
34	23	24	26	27	42	43	45	46
35	24	25	27	28	43	44	46	47
36	25	26	28	29	44	45	47	48
37	26	27	29	30	45	46	49	50
38	27	28	30	31	46	47	49	50
39	28	29	30	32	47	49	50	51
40	29	30	31	33	48	50	51	52
41	29	31	32	34	49	51	52	54
42	30	31	33	35	50	52	54	55
43	31	32	34	35	52	53	55	56
44	32	33	35	36	53	54	56	57
45	33	34	36	37	54	55	57	58
46	34	35	37	38	55	56	58	59
47	35	36	38	39	56	57	59	60
48	35	37	38	40	57	59	60	62
49	36	38	39	41	58	60	61	63
50	37	38	40	42	59	61	63	64

Tabla 13 Papel probabilístico normal



Formulario

Cuadro F.1 Análisis descriptivo de datos

	Datos sin agrupar	Datos agrupados
Frecuencia relativa [$fr(x_j)$]	$1/n$	(n.º de datos con x_j)/n
Media ($\overline{x}$)	$\Sigma x_i/n$	$\Sigma x_j fr(x_j)$
Desviación típica (s)	$\sqrt{\Sigma(x_i-\overline{x})^2/n}$	$\sqrt{\Sigma(x_j-\overline{x})^2 fr(x_j)}$
Desigualdad de Tchebychev	$fr(\lvert x_i-\overline{x}\rvert \leq k) \geq 1-1/k^2$	
Coeficiente de variación (CV)	$s/\lvert\overline{x}\rvert$	
Coef. de asimetría (CA)	$\Sigma(x_i-\overline{x})^3/ns^3$	$\Sigma\left(\frac{x_j-\overline{x}}{s}\right)^3 fr(x_j)$
Coef. de apuntamiento (CA_p)	$\Sigma(x_i-\overline{x})^4/ns^4$	$\Sigma\left(\frac{x_i-\overline{x}}{s}\right)^4 fr(x_j)$
Diagrama de caja	$L(S/I) = Q_1 \pm 1{,}5\,(Q_3-Q_1)$	
Transformaciones [$y=h(x)$]	$\overline{y} = h(\overline{x}) + h''(\overline{x})\,s_x^2/2 \quad ; \quad s_y = s_x\lvert h'(\overline{x})\rvert$	
Transformaciones ($y=a+bx$)	$\overline{y} = a + b\overline{x} \quad ; \quad s_y = \lvert b\rvert s_x$	
Relaciones entre frecuencias conjuntas, marginales y condicionadas	$fr(x_i y_i) = fr(x_i\lvert y_i) fr(y_i)$	
Covarianza	$\mathrm{Cov}(x,y) = \frac{\Sigma(x_i-\overline{x})(y_i-\overline{y})}{n}$	
Coeficiente de correlación	$r = \frac{\mathrm{Cov}(x,y)}{s_x s_y}$	
Coeficiente de regresión	$b = \frac{\mathrm{Cov}(x,y)}{s_x^2}$	
Desviación típica residual	$s_R = \sqrt{\frac{\Sigma(y_i-a-bx_i)^2}{n}}$	

Cuadro F.2 Probabilidad y distribuciones continuas

Para todo suceso A	$0 \le P(A) \le 1$
Si E es el suceso seguro	$P(E) = 1$
A y B mutuamente excluyentes	$P(A + B) = P(A) + P(B)$
$\bar{A}$ complementario de A	$P(\bar{A}) = 1 - P(A)$
Probabilidad conjunta	$P(AB) = P(A\|B)P(B)$
Independencia	$P(AB) = P(A)P(B)$
Teorema de Bayes	$P(A_i \mid B_j) = \dfrac{P(B_j \mid A_i)P(A_i)}{\sum_j P(B_j \mid A_i)P(A_i)}$
Función de distribución	$F(x_0) = P(x \le x_0)$
Función de densidad	$P(x_0 < x \le x_1) = \int_{x_0}^{x_1} f(x)\,dx$
Esperanza	$E(x) = \int_{-\infty}^{\infty} xf(x)dx = \mu$
Varianza	$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx$
Momento de orden k respecto a la media	$\mu_k = \int (x - \mu)^k f(x)dx$
Coeficiente de asimetría	$CA = \mu_3/\sigma^3$
Coeficiente de apuntamiento	$CA_p = \mu_4/\sigma^4$
Coeficiente de variación	$CV = \sigma/\mu$
Transformaciones, función de densidad	$g(y) = f(x)\left\lvert \dfrac{dx}{dy} \right\rvert$
Transformaciones, esperanza	$E[h(x)] = \int h(x)f(x)dx$

Cuadro F.3 Distribuciones de probabilidad para una variable

Nombre	Función de probabilidades o de densidad	Media	Varianza	C. asimetría	C. apuntamiento
Binomial	$p(x = r) = \binom{n}{r} p^r (1-p)^{n-r}$ $(r = 0, 1, ..., n)$	np	npq	$\frac{q-p}{\sqrt{npq}}$	$3 + \frac{1-6pq}{\sqrt{npq}}$
Geométrica	$p(x = r) = pq^{r-1}$ $(r = 1, 2, ...)$	$\frac{1}{p}$	$\frac{q}{p^2}$	$\frac{1+q}{\sqrt{q}}$	$3 + \frac{p^2+6q}{q}$
Poisson	$p(x = r) = \frac{\lambda^r e^{-\lambda}}{r!}$ $(r = 0, 1, ...)$	λ	λ	$1/\sqrt{\lambda}$	$3 + 1/\lambda$
Exponencial	$f(x) = \lambda e^{-\lambda x}$ $x > 0$	$1/\lambda$	$1/\lambda^2$	2	9
Normal	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-[(x-\mu)/\sigma]^2/2}$ $-\infty < x < \infty$	μ	σ^2	0	3
Uniforme	$f(x) = \frac{1}{b-a}$ $(a < x < b)$	$\frac{b+a}{2}$	$\frac{1}{12}(b-a)^2$	0	1,8
log normal	$f(x) = \frac{1}{x\sigma} \frac{1}{\sqrt{2\pi}} e^{-[(\ln x-\mu)/\sigma]^2/2}$ $x > 0$	$\exp\left(\mu + \frac{\sigma^2}{2}\right)$	$e^{\sigma^2}(e^{\sigma^2}-1)e^{2\mu}$	$(e^{\sigma^2}+2)\sqrt{e^{\sigma^2}-1}$	—
t con n g. de l.	$f(x) = k\left(1 + \frac{x^2}{n}\right)^{-[(n+1)/2]}$ $-\infty < x < \infty$	0	$\frac{n}{n-2}$ $(n > 2)$	0	$3 + \frac{6}{n-4}$ $(n > 4)$
χ^2 con n g. de l.	$f(x) = k(x^2)^{n/2-1} e^{-x^2/2}$ $x^2 \geq 0$	n	$2n$	$\sqrt{8/n}$	$3 + 12/n$
F con n_1, n_2, g. de l.	$f(x) = kx^{n_1/2-1}$ $(n_2 + n_1 x)^{(n_1+n_2)/2}$ $x > 0$	$\frac{n_2}{n_2-2}$ $(n_2 > 2)$	$\frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}$ $n_2 > 4$	—	—

Cuadro F.4 Distribuciones para varias variables

Distribución condicionada	$f(y\|x) = f(yx)/f(x) \quad [f(x) \neq 0]$
Independencia	$f(xy) = f(x)f(y)$
Momentos de sumas y diferencias	$E[x \pm y] = E[x] \pm E[y]$ $\text{Var}[x \pm y] = \text{Var}(x) + \text{Var}(y) \pm 2\ \text{Cov}(x, y)$
Matriz de covarianzas	$M = E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})']$
Estandarización en normal n-dimensional	$\mathbf{z} = \mathbf{A}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \quad , \quad \mathbf{M} = \mathbf{AA}'$ $f(\mathbf{z}) = \dfrac{1}{(2\pi)^{n/2}} \exp\left\{-\dfrac{1}{2}\mathbf{z}'\mathbf{z}\right\}$
Distancia de Mahalanobis	$D(x, \boldsymbol{\mu}) = (x - \boldsymbol{\mu})'\mathbf{M}^{-1}(x - \boldsymbol{\mu})$
Momentos de transformaciones lineales	$\mathbf{Y} = \mathbf{AX} \quad , \quad \mathbf{M}_y = \mathbf{AM}_x\mathbf{A}' \quad , \quad \boldsymbol{\mu}_y = \mathbf{A}\boldsymbol{\mu}_x$
Distribución multinomial	$P(n_1, ..., n_k) = \dfrac{n!}{n_1! \ldots n_k!} p_1^{n_1} \ldots p_k^{n_k}$
Distribución normal n-dimensional	$f(\mathbf{x}) = \dfrac{1}{\lvert\mathbf{M}\rvert^{1/2}(2\pi)^{n/2}} \exp\left\{-\dfrac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\mathbf{M}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\}$
ji-cuadrado	$\chi_n^2 = z_1^2 + ... + z_n^2 \quad ; \quad z_i \sim N(0, 1)$ indepen.
t de Student	$t_n = \dfrac{z_1}{\sqrt{\dfrac{\chi_n^2}{n}}}$
F de Fisher	$F_{n, m} = \dfrac{\chi_n^2/n}{\chi_m^2/m}$

Cuadro F.5 Estimación

Muestra aleatoria simple	$f(x_1, ..., x_n) = f(x_1) \dots f(x_n)$
Distribución media muestral	$E(\bar{x}) = \mu \quad ; \quad DT(\bar{x}) = \sigma/\sqrt{n}$
Corrección poblaciones finitas	$DT(\bar{x}) = (\sigma/\sqrt{n})\sqrt{(N-n)/(N-1)}$
Varianza muestral corregida	$\hat{s}^2 = \Sigma(x_i - \bar{x})/(n-1) \quad ; \quad E(\hat{s}^2) = \sigma^2$
Distribución varianza muestral	$(n-1)\hat{s}^2/\sigma^2 \sim \chi^2_{n-1}$
Estimador centrado	$E(\hat{\vartheta}) = \vartheta$
Sesgo o eficiencia	sesgo $(\hat{\vartheta}) = E(\hat{\vartheta}) - \vartheta$
Precisión o eficiencia	$ef(\hat{\vartheta}) = \text{Var}(\hat{\vartheta})^{-1}$
Eficiencia relativa	$ER(\hat{\vartheta}_2/\hat{\vartheta}_1) = \text{Var}(\hat{\vartheta}_1)/\text{Var}(\hat{\vartheta}_2)$
Error cuadrático medio	$E(\hat{\vartheta} - \vartheta)^2 = \text{sesgo}\,(\hat{\vartheta})^2 + \text{Var}(\hat{\vartheta})^2$
Combinación de estimadores centrados	$\hat{\vartheta}_T + \Sigma\alpha_i\hat{\vartheta}_i \quad ; \quad \alpha_i = ef(\hat{\vartheta}_i)/\Sigma_j ef(\hat{\vartheta}_j)$
Consistencia	$E(\hat{\vartheta}_n) \to \vartheta \quad ; \quad \text{Var}(\hat{\vartheta}_n) \to 0$
Estimador *MV* (máxima verosimilitud)	$\max_{\vartheta} L(\vartheta) = \max_{\vartheta} \ln f(\mathbf{x}\|\vartheta)$
Varianza asintótica estimador *MV*	$\hat{\sigma}^2_{MV} = \left[-\dfrac{d^2 L(\hat{\vartheta}_{MV})}{d\vartheta^2} \right]^{-1}$
Distribuciones en el muestreo	
$\hat{p}$ en binomial (p, n)	aprox. normal $(p, \sqrt{pq/n})$
$\bar{x}$, población cualquiera	aprox. normal $(\mu, \sigma/\sqrt{n})$
$\bar{x}$ en $N(\mu, \sigma)$	$N(\mu, \sigma/\sqrt{n})$
s^2 en $N(\mu, \sigma)$	$ns^2/\sigma^2 = (n-1)\hat{s}^2/\sigma^2 \sim \chi^2_{n-1}$

Cuadro F.6 Intervalos de confianza

a) *Proporciones*

Parámetro	Estadístico pivote	Distribución	Intervalo de confianza
p	$(\hat{p} - p)/\sqrt{\hat{p}\hat{q}/n}$	$N(0, 1)$	$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$
$p_1 - p_2$	$\frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}}$	$N(0, 1)$	$\hat{p}_1 - \hat{p}_2 \pm \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$

b) *Poblaciones normales*

Parámetro	Estadístico pivote	Distribución	Intervalo de confianza
μ, σ conocido	$\sqrt{n}(\bar{x} - \mu)/\sigma$	$N(0, 1)$	$\bar{x} \pm z_{\alpha/2}\sigma\sqrt{n}$
μ, σ desconocido	$\sqrt{n}(\bar{x} - \mu)/\hat{s}$	t_{n-1}	$\bar{x} \pm t_{\alpha/2}\hat{s}\sqrt{n}$
σ^2	$(n-1)\hat{s}^2/\sigma^2$	χ^2_{n-1}	$(n-1)\hat{s}^2/\chi^2_{\alpha/2}$, $(n-1)\hat{s}^2/\chi^2_{1-\alpha/2}$
$\mu_1 - \mu_2$ $\sigma_1 = \sigma_2$	$\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\hat{s}_T\sqrt{n_1^{-1} + n_2^{-1}}}$	$t_{n_1+n_2-2}$	$\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2}\hat{s}_T\sqrt{n_1^{-1} + n_2^{-1}}$
$\mu_1 - \mu_2$ $\sigma_1 \neq \sigma_2$	$\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\hat{s}_1^2/n_1 + \hat{s}_2^2/n_2}}$	$t_{n_1+n_2-\Delta-2}$	$\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2}\sqrt{\hat{s}_1^2/n_1 + \hat{s}_2^2/n_2}$

c) *Intervalos asintóticos*

Parámetro	Estadístico pivote	Distribución	Intervalo de confianza
ϑ	$\frac{\vartheta - \hat{\vartheta}_{MV}}{\sigma(\hat{\vartheta}_{MV})}$	$N(0, 1)$	$\hat{\vartheta}_{MV} \pm z_{\alpha/2}\sigma(\hat{\vartheta}_{MV})$

Cuadro F.7 Contrastes de hipótesis

Población	Parámetro	Contraste	Región de aceptación
a) Contraste para una población			
Binomial	p	H_0: $p = p_0$ H_1: $p \neq p_1$	$\lvert\hat{p} - p_0\rvert < z_{\alpha/2}\sqrt{\dfrac{p_0 q_0}{n}}$
Normal o muestras grandes	μ	H_0: $\mu = \mu_0$ H_1: $\mu \neq \mu_0$	$\lvert x - \mu_0\rvert \leq t_{\alpha/2}\dfrac{\hat{s}}{\sqrt{n}}$
Normal	σ^2	H_0: $\sigma^2 = \sigma_0^2$ H_1: $\sigma^2 > \sigma_0^2$	$\dfrac{(n-1)\hat{s}^2}{\sigma_0^2} \leq \chi^2\,(n-1;\,\alpha)$
Cualquiera	$\underline{\vartheta}$	H_0: $\underline{\vartheta} \in \Omega_0$ H_1: $\underline{\vartheta} \in \Omega - \Omega_0$	$2 \ln \lambda = 2[L(\hat{\Omega}) - L(\hat{\Omega}_0)] < \chi_g^2$
b) Contrastes para dos poblaciones			
Binomiales	p_1, p_2	H_0: $p_1 = p_2$ H_1: $p_1 \neq p_2$	$\lvert\hat{p}_1 - \hat{p}_2\rvert < z_{\alpha/2}\sqrt{\hat{p}_0\hat{q}_0\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}$ $\hat{p}_0 = \alpha\hat{p}_1 + (1-\alpha)\hat{p}_2$ $\alpha = n_1/(n_1 + n_2)$
Normales con misma σ^2	μ_1, μ_2	H_0: $\mu_1 = \mu_2$ H_1: $\mu_1 \neq \mu_2$	$\lvert\bar{x}_1 - \bar{x}_2\rvert \leq z_{\alpha/2}\,\hat{s}_T\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$ $\hat{s}_T = \sqrt{\alpha s_1^2 + (1-\alpha)\hat{s}_2^2}$ $\alpha = (n_1 - 1)/(n_1 + n_2 - 2)$
Normales apareadas	μ_1, μ_2	H_0: $\mu_1 = \mu_2$ H_1: $\mu_1 \neq \mu_2$	$\lvert\bar{y}\rvert \leq t_{\alpha/2}\dfrac{\hat{s}_y}{\sqrt{n}}$ $y = x_1 - x_2$
Normales	σ_1^2, σ_2^2	H_0: $\sigma_1^2 = \sigma_2^2$ H_1: $\sigma_1^2 > \sigma_2^2$	$\hat{s}_1^2/\hat{s}_2^2 \leq F(n_1 - 1, n_2 - 1;\,\alpha)$
Normales con distinta σ^2	μ_1, μ_2	H_0: $\mu_1 = \mu_2$ H_1: $\mu_1 \neq \mu_2$	$\lvert\bar{x}_1 - \bar{x}_2\rvert \leq t(k)s(k)\sqrt{\dfrac{1}{n_1} + \dfrac{k^2}{n_2}}$ $k = \sigma_2/\sigma_1$

Cuadro F.8 Contrastes de ajuste

1. Contrastes de ajuste	Nombre	Estadístico	Distribución
General	*ji*-cuadrado	$\Sigma \frac{(\text{Observadas-Esperadas})^2}{\text{Esperadas}}$	χ^2_{n-p-1}
V. continuas	Kolmogorov-Smirnov	Sup $\lvert F_n(x) - F(x)\rvert$	Tabla 8
Normalidad	Shapiro y Wilk	$\frac{1}{ns^2} \sum_{1}^{n} a_{j,n}[x_{(n-j-1)} - x_{(j)}]$	Tablas 10 y 11
	K-S-Lilliefors	Sup $\lvert F_n(x) - F(x)\rvert$	Tabla 9
	Asimetría	$\sqrt{n}\, CA/\sqrt{6}$	$N(0, 1)$
	Apuntamiento	$\sqrt{n}\,(CAp - 3)/\sqrt{24}$	$N(0, 1)$
	Conjunto	$n[CA^2/6 + (CAp - 3)^2/24]$	χ^2_2
	Transformación de Box-Cox	$x^{(\lambda)} = \frac{(x+m)^{\lambda} - 1}{\lambda}$	
Estimación de densidades	Métodos núcleo	$f(x) = \frac{1}{hn} \Sigma \frac{1}{\sqrt{2\pi}} e^{-1/2(x-x_i)^2/h^2}$	

Cuadro F.9 Estadísticas para la validación del modelo

2. Independencia	Rachas	x = n.º de rachas	$x \sim N\left[k+1; \dfrac{k(k+1)}{2k-1}\right]$ y tabla 12
	Autocorrelación	$Q = n(n+2)\sum_{1}^{m} \dfrac{r^2(k)}{n-k}$	χ^2_{m-1}

3. Homogeneidad	Nombre	Estadístico	Distribución
Dos muestras	Wilcoxon	Rango de los datos de una muestra	$R_x \sim N\left[n\dfrac{(N+1)}{2}; \sqrt{\dfrac{nm}{12}(N+1)}\right]$
Atributos	Tablas de contingencia ($r \times c$)	$\Sigma \dfrac{(\text{Observadas-Esperadas})^2}{\text{Esperadas}}$	$\chi^2(r-1)\times(c-1)$
Un dato atípico población normal	Desviación máxima estudentizada	máx $\left\lvert \dfrac{x_i - \bar{x}}{\hat{s}} \right\rvert$	Tabla 6.8
Varios datos atípicos poblaciones normales	Coeficiente de apuntamiento	$\sqrt{n}(CAp-3)/\sqrt{24}$	Tablas 6.9 y $N(0, 1)$

Resolución de ejercicios

Capítulo 2

Ejercicios 2

2.1. a) $\bar{x} = (28 + 22 + ... + 35)/16 = 38{,}88$; Med $= (38 + 37)/2 = 37{,}5$;
$s_x^2 = [(28 - 38{,}88)^2 ... (35 - 38{,}88)^2]/16 = 114{,}73$; $s_x = 10{,}71$.
b) $\bar{x} = 25(3/16) + ... + 65(1/16) = 40$; Med $= 35$;
$s_x^2 = [(25 - 40)^2(3/16) ... (65 - 40)^2(1/16)] = 125$; $s_x = 11{,}18$.
c)

2	825
3	521875
4	2410
5	38
6	1

2.6. $M = \Sigma(x_i - a)^2$; $dM/da = 0 = 2\Sigma(x_i - a)$; $a = \Sigma x_i/n = \bar{x}$.
La media minimiza el error cuadrático medio.

2.7. $y = kx$; $\bar{y} = k\Sigma x/n = k\bar{x}$; $s_y^2 = \Sigma(kx - k\bar{x})^2/n = k^2 s_x^2$.

2.8. Las medias de longitud son 5,685 y 4,525; las medianas, 4,87 y 4,25; y las desviaciones típicas, 1,519 y 1,067. Los coeficientes de asimetría y curtosis son 0,59 y – 0,11 para asimetría y 2,4 y 2,3 para curtosis. En el primer CD las canciones son más largas y con mayor variabilidad en las longitudes.

2.9. $\bar{z} = \Sigma z_i/n = \Sigma(x_i + y_i)/n = \bar{x} + \bar{y}$.

2.10. $\bar{z} = \left(\sum_{n_1} x_i + \sum_{n_2} y_i\right)/(n_1 + n_2) = (n_1\bar{x} + n_2\bar{y})/(n_1 + n_2)$.

2.11. $s_z^2 = \Sigma(z - \bar{z})^2/n = \Sigma(x + y - \bar{x} - \bar{y})^2/n = \Sigma(x - \bar{x})^2/n + \Sigma(y - \bar{y})^2/n + 2\Sigma(x - \bar{x}) \cdot (y - \bar{y})/n = s_x^2 + s_y^2 + 2\Sigma(x - \bar{x})(y - \bar{y})/n$. La varianza de la suma puede ser mayor o menor que la suma de varianzas, dependiendo del signo de $\Sigma(x - \bar{x})(y - \bar{y})$.

2.12. $\mu_3 = \Sigma(x_i - \bar{x})^3/n = \Sigma(x_i^3 - 3x_i^2\bar{x} + 3x_i\bar{x}^2 - \bar{x}^3)/n = m_3 - 3m_2m_1 + 3m_1^3 - m_1^3$.

2.13. a) $s_x^2 = (0{,}2 \cdot 2^2 + 0{,}2 \cdot 1^2)2 = 2$; $s_y^2 = a^2 0{,}1 + 0{,}004$; $a = 4{,}47$.
b) $Cap(x) = 6{,}8/4 = 1{,}70$; $Cap(y) = (0{,}1a^4 + 0{,}1^4 \cdot 0{,}4)/s_y^4$; para $a = 4{,}47$, $Cap(y) = 10$.

2.14. La varianza de la distribución.

2.15. b) Michelson $\bar{x} = 896{,}67$; Newcomb $\bar{x} = 763{,}20$.
c) Las medias de Michelson están desplazadas respecto a las de Newcomb, no parecen estar midiendo lo mismo.

2.16. a) $\ln G = 1/n\Sigma \ln x_i$; si $y = \ln x$, $\ln G = \bar{y}$;
b) si $y = x^{-1}$, $H = \bar{y}^{-1}$;
c) $\bar{y} = \ln\bar{x} - (1/2)(s_x^2/\bar{x}^2)$; $G = \bar{x}e^{-(1/2)(s_x^2/\bar{x}^2)}$; $\bar{x} > G$; análogamente: $\bar{x}/H = 1 + s_x^2/\bar{x}^2$.

2.17. Como $V_k/V_0 = V_k/V_{k-1} \cdot V_{k-1}/V_{k-2} \ldots \cdot \dfrac{V_1}{V_0} = r_k \ldots r_1$, si $(r_m)^k = r_1 \ldots r_k \Rightarrow r_m =$ media geométrica (r_i).

2.18. $\sum_i \sum_j (x_i - x_i)^2 = \Sigma\Sigma x_i^2 + \Sigma\Sigma x_j^2 - 2\ \Sigma x_i \Sigma x_j = 2n\ \Sigma x_i^2 - 2n^2\bar{x}^2 = 2n^2\left(\dfrac{\Sigma x_i^2}{n} - \bar{x}^2\right) = 2n^2 s^2$.

Capítulo 3

Ejercicios 3

3.1. $x' = k_1x$; $y' = k_2y$; $\text{Cov}(x', y') = 1/n\Sigma(x' - \bar{x}')(y' - \bar{y}') = k_1k_2\text{Cov}(x, y)$, como según ejercicio 2.7 las desviaciones típicas quedan multiplicadas por k_1 y k_2; $r(x'y') = r(x, y)$.

3.2. Si $y = a + bx$, $\text{Cov}(x, y) = 1/n\ \Sigma(x - \bar{x})(y - \bar{y}) = 1/n\ \Sigma(x - \bar{x})(a + bx - a - b\bar{x}) = = bs_x^2$, $s_y^2 = 1/n\ \Sigma(y - \bar{y})^2 = 1/n\ \Sigma(a + bx - a - b\bar{x})^2 = b^2s_x^2$. Por tanto, $r(x, y) = = bs_x^2/(s_x \cdot bs_x) = 1$.

3.3. La matriz de varianzas y covarianzas de dos variables es semidefinida positiva; por tanto, el determinante:

$$\begin{bmatrix} s_x^2 & \text{Cov}(x, y) \\ \text{Cov}(x, y) & s_y^2 \end{bmatrix} \geq 0, \text{ es decir: } s_x^2 s_y^2 - \text{Cov}^2(x, y) \geq 0;$$

$$1 \geq \frac{\text{Cov}^2(x, y)}{s_x^2 s_y^2} \text{ que implica } 1 \geq r^2, \text{ es decir, } -1\ r \leq 1.$$

3.4. Las medias son 2 y 8 y la covarianza será

$\text{Cov}(x, y) = \Sigma xy/n - \overline{xy} = (16 + 9 + 16 + 5 + 40)/5 - 16 = 1{,}2$. Las desviaciones típicas serán
$s_x = \sqrt{(0 + 1 + 0 + 1 + 4)/5} = \sqrt{1{,}2} = 1{,}0954$ y $s_y = \sqrt{(0 + 1 + 0 + 9 + 4)/5} =$
$= \sqrt{2{,}8} = 1{,}67$. El coeficiente de correlación será $r = 1{,}2/(1{,}0954 \times 1{,}67) = 0{,}65$.

3.5. La recta de regresión es $y = 8 + b(x - 2)$ y la pendiente $b = \text{Cov}(x, y)/s_x^2 =$
$= 1{,}2/1{,}2 = 1$.

3.6. Si intercambiamos x por y la nueva recta será $x = 2 + c(y - 8)$ donde $c = \text{Cov}(x, y)/s_y^2 = 1{,}2/2{,}8 = 0{,}429$, es decir, $x = 2 + 0{,}429\ (y - 8)$. El producto de las pendientes es 0,429, que es $0{,}65^2$. Esto ocurrirá siempre, ya que el producto de las dos pendientes es $[\text{Cov}(x, y)/s_x^2] \times [\text{Cov}(x, y)/s_y^2] = \text{Cov}\ (x, y)^2/s_x^2 s_y^2 = r^2$.

3.7. Supongamos, sin pérdida de generalidad, que ordenamos las parejas de manera que las primeras n_1 tienen -1 de la variable cualitativa y las siguientes n_2 tienen $+1$. La media de la variable cualitativa será $(n_2 - n_1)/n$, y llamando m_1 a la media de los primeros n_1 datos de y y m_2 a la media de los siguientes n_2 la media de esta variable es $(n_1 m_1 + n_2 m_2)/n$. La covarianza será $\text{Cov}(x, y) = \Sigma xy/n - \overline{xy} = (-\Sigma_{n_1}$
$y + \Sigma_{n_2} y)/n - [(n_1 m_1 + n_2 m_2)/n] \times [(n_2 - n_1)/n] = (-n_1 m_1 + n_2 m_2)/n - (n_1 m_1 +$
$+ n_2 m_2)(n_2 - n_1)/n^2 = [(n_1 + n_2)(-n_1 m_1 + n_2 m_2) - (n_1 m_1 + n_2 m_2)(n_2 - n_1)]/n^2 =$
$= 2n_1 n_2 (m_1 - m_2)/n^2$.

3.10. Si hacemos un gráfico entre las velocidades y las distancias se observa claramente una relación lineal. La recta de regresión es *vel* = 0,0746 + 0,0353 dis. Por ejemplo, la velocidad estimada para una galaxia situada a una distancia de 1.000 millones de años luz es *vel* = 0,0746 + 0,0353 × (1.000) = 0,0746 + 35,3 = 35,37 miles de millas por segundo.

3.11. El gráfico indica una relación aproximadamente lineal y la recta de regresión es *propor* = 54,61 – 1,16 *renta*.

Capítulo 4

Ejercicios 4.1

4.1.1. Prob. de 1.ª bola par 2/5; segunda impar 3/4; primera impar 3/5; segunda par 2/4. $p = (2/5) \cdot (3/4) + (3/5) \cdot (2/4) = 6/10$. También: casos favorables $\binom{2}{1}\binom{3}{1} = 6$; casos posibles $\binom{5}{2} = 10$; $p = 6/10$.

4.1.2. $P(D) = P(M_1D) + P(M_2D) + P(M_3D) = P(M_1)P(D|M_1) + P(M_2)P(D|M_2) +$

$$+ P(M_3)P(D|M_3) = \frac{300}{1.350}\,0{,}02 + \frac{450}{1.350} \cdot 0{,}035 + \frac{600}{1.350} \cdot 0{,}025 = 0{,}0272.$$

4.1.3. $P(A) = P(B) = P(C) = 0{,}5$. $P(AB) = P(AC) = P(BC) = 0{,}25$ y verifica la condición de independencia dos a dos, pero $P(ABC) = 0{,}25 \neq 0{,}5^3$ y por tanto no son independientes.

4.1.4. $P(R|V) = 0{,}3$; $P(R|M) = 0{,}1$; $P(V) = 0{,}6$. Entonces, $P(MR) = P(R|M)P(M) = 0{,}1 \cdot 0{,}04 = 0{,}04$; $P(VR) = P(R|V)P(V) = 0{,}3 \cdot 0{,}6 = 0{,}18$.

$P(R) = P(MR) = P(VR) = 0{,}04 + 0{,}18 = 0{,}22$.

$$P(M|R) = \frac{P(MR)}{P(R)} = \frac{0{,}04}{0{,}22} = 0{,}18.$$

4.1.5. $$P(M_1M_1|BB) = \frac{P(BB|M_1M_1)P(M_1M_1)}{P(BB)} = \frac{0{,}95^2 \cdot (1/4)}{0{,}8326} = 0{,}2710,$$ donde

$P(BB) = \Sigma P(BB y M_iM_j) = \Sigma P(M_iM_j)P(BB|M_iM_j) = 0{,}8326$.

$$P(M_2M_2|BB) = \frac{0{,}9^2 \cdot (1/16)}{0{,}8326} = 0{,}0608;\ P(M_3M_3)|BB) = \frac{0{,}85^2 \cdot (1/16)}{0{,}8326} = 0{,}0542;$$

$P_{(\text{pedida})} = 0{,}2710 + 0{,}0608 + 0{,}0542 = 0{,}386$.

4.1.7. Si son mutuamente excluyentes $P(AB) = 0$. Por tanto $P(AB) \neq P(A)P(B)$ y son dependientes.

4.1.8. a) $0{,}5 \cdot 0{,}25^2$.

b) $\frac{2}{5}\,0{,}5 + \frac{2}{5}\,0{,}75 + \frac{1}{5}\,0{,}75$.

c) $\left(\frac{2}{5}\right)^3 + \left(\frac{2}{5}\right)^3 + \left(\frac{1}{5}\right)^3$.

d) $3!(4/125)$.

4.1.9. Suponemos que los días son equiprobables y los sucesos independientes:
$$1 - \frac{364}{365} \cdot \frac{363}{365} \cdot \frac{362}{365} \ldots \frac{365 - N + 1}{365}.$$

4.1.10. 1) $1 - 0{,}99^{50} = 0{,}39$.
2) $1 - (0{,}99^{50} + 50 \cdot 0{,}99^{49} \cdot 0{,}01)$.
3) $0{,}01^{50}$.

4.1.11. $P(D/\text{Muestra}) = (0{,}09^{10} \cdot 0{,}3)/(0{,}9^{10} \cdot 0{,}3 + 0{,}99^{10} \cdot 0{,}7) = 0{,}14$.

4.1.12. $0{,}5 = 1 - (364/365)^n$; $n = 253$.

4.1.13. Sea p_A la prob. de ganar a A y p_B a B. $p_A < p_B$. Entonces, la secuencia ABA ofrece probabilidad $p_A p_B + p_A p_B(1 - p_A) = p_A p_B(2 - p_A)$ y la BAB $p_A p_B(2 - p_B)$. Como $p_B > p_A$, la secuencia ABA será preferida.

4.1.14. Probabilidad de veredicto correcto el jurado = p(mayoría) + p(unanimidad) = $= pq + p^2/2 + p^2/2 = p^2 + p(1 - p) = p$ = juez individual. Ambos la misma.

4.1.15. $p^3 + 3p^2q > p$; para $p > 0{,}5$ es mejor el jurado.

4.1.16. El 7 primero, el 3 después. Haga la prueba.

4.1.17. a) $1 - P(\text{todas distintas}) = 1 - \left(\frac{52}{52} \cdot \frac{48}{51} \cdot \frac{44}{50} \cdot \frac{40}{49} \cdot \frac{36}{48}\right) = 0{,}49$.

b) $\binom{13}{2}\binom{4}{2}\binom{4}{2}\binom{44}{1} \Big/ \binom{52}{5}$

c) $\binom{13}{1}\binom{4}{3}\binom{48}{2} \Big/ \binom{52}{5} + \binom{13}{1}\binom{4}{4}\binom{48}{1} \Big/ \binom{52}{5} = \binom{13}{1}\binom{4}{3}\binom{49}{2} \Big/ \binom{52}{5}$

d) $2\binom{13}{2}\binom{4}{2}\binom{4}{3} \Big/ \binom{52}{5}$.

e) $\binom{13}{1}\binom{4}{4}\binom{48}{1} \Big/ \binom{52}{5}$.

Ejercicios 4.2

4.2.2. $k = (b - a)^{-1}$.

4.2.3. Mediana $= a + (b - a)/2 = (a + b)/2$.

4.2.4. a) nx^{n-1}.
b) $\ln x = (1/n)\ln 0{,}5$.
c) $E(x) = n/(n+1)$; Var $(x) = n/(n+2)(n+1)^2$.

4.2.5. a) x^3.
b) $(2/3)^3$; $(9/10)^3$; $(1/2)^3 - (1/3)^3$.
c) 0,63.
d) $\overline{x} = 0{,}75$; Var$(x) = 3/80$.

4.2.6. a) $k = 3/4$.
b) Si $y = 100x$, $f(y) = 100^{-3}(3/4)(y - 100)(300 - y)$ para $(100 \leq y \leq 300)$.
c) En metros $y = 2x$, $f(y) = (3/4)2^{-3}(y - 2)(6 - y)$; $2 \leq y \leq 6$.
d) $y = \pi x^2$, $f(y) = (3/4) \cdot (2\sqrt{\pi y})^{-1}(\sqrt{y/\pi} - 1)(3 - \sqrt{y/\pi})$
e) $1 - \int_{1,2}^{2,8} f(x)dx$.

4.2.7. Apostando x; $(18/37)x + (19/37)(-x) = -x \cdot 0{,}027$. Perderemos el 2,7% de lo jugado, a largo plazo.

4.2.8. $0{,}001(-5.000.000) + 0{,}999(0) + 10.000 = 5.000$ ptas. de beneficio esperado.

4.2.9. $F(x)$ debe ser siempre positiva, luego para $x \leq a$, $F(x) = 0$; $x \geq b$, $F(b) = 1$; $f(x) = (b - a)^{-1}$ $(a \leq x \leq b)$.

4.2.10. b) $f(x) = 1/2$, $0 \leq x \leq 1$; $f(x) = 1/4$, $2 \leq x \leq 4$; cero en otro caso. Se trata de dos tipos de averías distintas que se producen con probabilidad 1/2 cada una. Reparar la primera requiere entre cero y 1 hora; la segunda entre 2 y 4.
c) $P(x > 3{,}5/x > 1) = 1{,}4$.

4.2.11. a) $1/2\sqrt{y}$, $0 \leq y \leq 1$.
b) $2y$, $0 \leq y \leq 1$.
c) $1/y^2$, $1 \leq y \leq \infty$.

4.2.12. a) $m = 1/4$.
b) $E(x) = 2$.
c) $F(x) = x^2/8$, $0 < x < 2$; $F(x) = x - 1 - x^2/8$, $2 < x < 4$.

4.2.13. $\int_{1,7}^{2,4} f(x)dx = 0{,}502$. Luego: $0{,}502^5 = 0{,}03$.

Capítulo 5

Ejercicios 5.1

5.1.1. $P(A) = \binom{10}{1}\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^9 = 0{,}32$; $P(B) = \binom{10}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^8 = 0{,}29$; más probable el A.

5.1.2. $P(\text{seis}) = 1/6$; $P(3 \text{ seises}) = \binom{6}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^3$. Más probable un seis.

5.1.3. $p_r/p_{r-1} = \binom{n}{r}p^r q^{n-r}/\binom{n}{r-1}p^{r-1}q^{n-r+1}$.

5.1.4. a) $e^{-3}3^6/6!$

b) $e^{-6}6^3/3!$

c) $1 - \sum_0^{15} e^{-9}9^i/i!$

d) $e^{-3/5}(3/5)^2/2!$

5.1.5. e^{-3}.

5.1.6. a) $(49.999/50.000)^{200.000} = 0{,}02$.

b) $\lambda = 200/50 = 4$; $P(\text{pedida}) = 1 - \sum_0^2 4^i e^{-4}/i! = 0{,}111$.

5.1.7. $1 - \sum_0^2 \binom{n}{i}0{,}01^i \cdot 0{,}99^{n-i} = 0{,}9$. Por Poisson $\sum_0^2 p(x = i) = 0{,}1$; con $\lambda = 5{,}2$, $p = 0{,}109$; con $\lambda = 5{,}4$, $p = 0{,}095$. Interpolando $\lambda = 5{,}3286$ y $n = \lambda/p = 533$.

5.1.8. x_i = n.º de circuitos que necesita el calcular i; $E(x_i) = 1{,}875$; $\text{Var}(x_i) = 1{,}107$, $y = \Sigma x_i$ será normal (18.875; 105,2); $p(x > 18.875) = p(z > 1{,}188) = 0{,}117$.

5.1.9. El número medio de paquetes necesario para el primer cupón es 1, para el segundo distinto 6/5 (tiene probabilidad 5/6), para el tercero 6/4 (tiene probabilidad 4/6), para el cuarto 6/3, para el quinto 6/2, para el sexto 6. Por tanto, 1 + + 6/5 + 6/4 + 6/3 + 6/2 + 6 = 14,7.

5.1.10. a) $p = 0{,}02$; $p(\text{aceptar lote}) = 0{,}98^{20} + 20 \cdot 0{,}98^{19} \cdot 0{,}02 = 0{,}94009$.

b) $p = 0{,}008$; $p(\text{rechazar}) = 1 - p(\text{aceptar}) = 1 - (0{,}8515 + 0{,}137) = 0{,}0115$.

5.1.11. $(5.800.000 - \mu)/\sigma = 2{,}33$; $(1.200.000 - \mu)/\sigma = -1{,}28$; $\mu = 2.831.020$;
$\sigma = 1.274.240$
$$p\left(x > \frac{3.000.000 - 2.831.020}{1.274.240}\right) = 0{,}55177.$$
$P(\text{pedida}) = 1 - 0{,}5517 = 0{,}4483.$

5.1.12. a) $p[x > (241{,}5 - 240)/3{,}098] = 0{,}3142$.
b) $p[x < (10{,}5 - 15)/3{,}814] = 0{,}119$.
c) $p[(5{,}5 - 10)/3{,}146 < x < (12{,}5 - 10)/3{,}146] = 0{,}7103$.

5.1.13. a) $1 - p(149{,}2 < x < 150{,}4) = 1 - p(-2 < x < 1) = 0{,}1815$.
b) $\binom{50}{44} 0{,}8185^{44} \cdot 0{,}1815^{6} = 0{,}08457$.

5.1.14. $E(y) = 2 - 1 = 1$; $\text{Var}(y) = 4E[(x^2 - 1)^2] = 4 \cdot 2 = 8$; $p(2x^2 - 1 - 1 < 2\sqrt{2}) =$
$= p(x^2 - 1 < \sqrt{2}) = p(-\sqrt{2{,}41} < x < \sqrt{2{,}41}) = 0{,}8788$.

5.1.15. a) $f(y) = \dfrac{1}{\ln 10 \cdot 0{,}8\sqrt{2\pi}} \dfrac{1}{y} e^{-(1/2)\cdot(\log y - 2)^2/0{,}82}$.
b) $p(y < 15) = p(x < \log 15) = p[z < (\log 15 - 2)/0{,}8] = 0{,}1515$; $p(15 < y <$
$< 4.000) = 0{,}8259$; $p(y > 4.000) = p(x > \log 4.000) = 0{,}0226$.

5.1.16. x = número de pasajeros que se presentan, x es normal con $\mu = 495 \cdot 0{,}88$ y $\sigma^2 =$
$= 495 \cdot 0{,}88 \cdot 0{,}12 \cdot p(x > 450) \simeq p(x \geqslant 450{,}5) = p\left(z \geqslant \dfrac{450{,}5 - 435{,}6}{7{,}22}\right) =$
$= p(z \geqslant 2{,}06) = 0{,}0197$.

5.1.17. Es una normal, con $\mu = 0$, $\sigma = 200$ truncada en σ, por tanto: $k = (0{,}1587 \cdot$
$\cdot 100\sqrt{2\pi})^{-1}$; p(avería antes de 250 h de un tubo) $= (0{,}8944 - 0{,}8413)/0{,}1587$.
$$p \text{ (al menos 65)} = p(y \leq 35{,}5) = p\left(\frac{y - 33{,}5}{4{,}72} \leqslant \frac{35{,}5 - 33{,}5}{4{,}72}\right) = p(z < 0{,}424) =$$
$= 0{,}6642$.

5.1.18. $\lambda = 3{,}87/7{,}5 = 0{,}52$; $p(x \geqslant 1) = 1 - p(x = 0) = 1 - e^{-0{,}52} = 0{,}4$.

5.1.19. a) $y = x + 1$, donde x es geométrica; por tanto $p(y = r) = p(x = r - 1) = q^{r-1} \cdot p$;
$r = 1, 2, \ldots$
b) $E[y] = E[x] + 1 = q/p + 1 = 1/p$.
c) Como $E[(y - E[y])^k] = E[(x - E[x])^k]$, tendrá la misma varianza y los mismos coeficientes de asimetría y apuntamiento.

5.1.20. a) Si z es binomial negativa, $p(z = r) = p(s = r + k)$, y $z = s - k$; $p(s = r) =$
$= p(z = r - k)$ para $s = k, k + 1, \ldots$; $p(s = r) = \binom{r-1}{r-k} p^k q^{r-k}$.
b) $E[s] = E[z] + k = kq/p + k = k/p$.
c) Como $E[(s - E[s])] = E[z + k - E(z) - k] = E[z - E(z)]$, tendrán los mismos coeficientes de asimetría y apuntamiento y la misma varianza.

5.1.21. a) Cuando p es aproximadamente 0 y $q \doteq 1$, entonces $\lambda = np \doteq npq$; para geométrica $q \doteq 1$, $1/p = 1/\lambda$; $1/p^2 = 1/\lambda^2$. Por tanto, deberían ser k/λ y k/λ^2 (véase el apéndice sobre la distribución gamma).

5.1.22. $y = x_1 + ... + x_n$, donde x_i toma valor 1 con probabilidad media $\bar{p}$ y con $\bar{q}$. $E(y) = \Sigma E(x_i) = n\bar{p}$; $\text{Var}(y) = \Sigma p_i q_i = \Sigma p_i - \Sigma p_i^2 = n\bar{p} - (n\sigma_p^2 + n\bar{p}^2) = n\bar{p}(1 - \bar{p}) - n\sigma_p^2$.

5.1.23. $E(y) = n\bar{p}$; $\text{Var}(y) = n\overline{pq} + n(n-1)n\sigma_p^2$.

Capítulo 6

Ejercicios 6

6.1. Ejemplo: $F(x, y) = 2/30$, $0 \le x \le 1$; $F(x, y) = 12/30$, $3 \le x \le 4$; $F(x, y) = 24/30$, $5 \le x \le 6$.

6.2. Si y tiempo mujer, x tiempo hombre, la condición es $x \le y + 15$ si $x \ge y$; $y \le x + 15$ si $x \le y$. Prob. = Área encuentro/Área total = 0,44.

6.3. P(pertenecer al cuadrado) $= (1 - 1/e)^2 \simeq 0{,}4$; P(pedida) $= 1 - (2/e - 1/e^2)^3 \simeq \simeq 0{,}78$.

6.4. $f(x_1, x_2) = [8(a-1)]^{-1}$; $P(x_1 > 4x_2^2) = [6(a-1)]^{-1}$, luego $a \ge 17{,}67$.

6.5. Marginales: $f(x) = 3x^2$, $0 < x < 1$; $f(y) = 3/2(1 - y^2)$, $0 < y < 1$. Condicionadas: $f(x/y) = 2x/(1 - y^2)$, $y < x < 1$; $f(y/x) = x^{-1}$ $(0 < y < x)$.

6.6. a) $E[x] = 4$.
b) $E[x - y] = 2$.
c) $\text{Cov}(x, y) = 4$; $\rho = 1/\sqrt{2} = 0{,}71$.

6.8. $E[z] = \mu\Sigma a_i$; $\text{Var}(z) = \sigma^2\Sigma a_i^2$; $\text{Cov}(y_i, z) = \sigma^2 a_i$; $r = a_i/\sqrt{\Sigma a_j^2}$

6.9. a) $Z = X + Y$ es normal con $\mu = 81{,}5 + 14{,}5 = 96$; $\sigma = (8^2 + 6^2)^{1/2} = 10$.
b) Normal bivariante con $\rho = 0$.

6.10. a) $v = x + \sum_1^{40} z_i$; con $x \sim N(520; 50)$, $z_i \sim N(96; 10)$; $E(v) = 4{,}36$ kg.; Desv. tipo$(v) = 80{,}6$ gr.

b) $p(x - \sum_1^5 z_i < 0) = P(H < 0)$, donde $H \sim N(40; 54, 77)$, $P(H < 0) = 0{,}2327$.

6.11. $P(T > C) = P(T - C > 0)$, $Y = T - C$ es normal $(-40; \sqrt{500})$, $P(Y > 0) = P(z > > 1{,}79) = 0{,}0367$.

6.12. $\text{Var}(z_1) = 92$; $\text{Var}(z_2) = 2/3$; $E(z_1 z_2) = 22/3$; $\rho = \dfrac{11}{\sqrt{138}}$.

Capítulo 7

Ejercicios 7.1

7.1.1. a) Como $F(x) = (x - 10)/40$; $10 < x < 20$; $x = 10\ NA + 10$, siendo NA el número aleatorio.
b) $F(x) = 1 - e^{-\lambda x}$; $x = -\ln(1 + NA)/\lambda$.

7.1.2. Y será asintóticamente $N(0,1)$.

Ejercicios 7.2

7.2.3. En la población $E(x) = \sum_{1}^{N} i \dfrac{1}{N} = \dfrac{(N+1)}{2}$. Por tanto $\hat{N} = 2\,\bar{x} - 1$.

7.2.4. $\hat{n} = x/p$.

7.2.5. $E(x) = a\sqrt{2\pi}$; por tanto $\hat{a} = \bar{x}/\sqrt{2\pi}$.

7.2.7. a) $E(x) = a/3$; $\hat{a} = 3\bar{x}$.
b) $E(x) = a\,x_0/(a-1)$; $\hat{a} = \bar{x}/(\bar{x} - x_0)$.

Ejercicios 7.3

7.3.1. $E(\Sigma\, \lambda_i \hat{\theta}) = \Sigma\, \lambda_i E(\hat{\theta}_i) = \theta\, \Sigma\, \lambda_i = \theta$.

7.3.2. $M = E[(\mu - a\bar{x})^2] = \mu^2 + a^2 \left(\dfrac{\sigma^2}{n} + \mu^2 \right) + 2a\mu^2$; $a = \mu^2/(\mu^2 + \sigma^2/n)$.
Obsérvese que $a < 1$.

7.3.3. $\text{Var}(\bar{x}) \to 0$ cuando $n \to \infty$ y es centrado.

7.3.4. $\hat{p} = x/n$ es centrado por p. $\text{Var}(\hat{p}) = \sigma^2/n = pq/n$. Es consistente.

7.3.5. $\hat{\lambda} = \bar{x} = 2{,}86$; Var $(\hat{\lambda}) = \sigma^2/n = 2{,}86/7 = 0{,}41$.

7.3.6. Var $[(x_1 + x_2)/2] = \sigma^2/2$ y no tiende a cero con n.

Ejercicios 7.4

7.4.1. $L = p_A^2 \cdot p_B^3 \cdot p_C^{25} = (2p_B)^2 p_B^3 (1 - 3\,p_B)^{25}$; $\dfrac{\partial L}{\partial p_B} = 0$ implica $p_B = \dfrac{1}{18}$;

$p_A = \dfrac{2}{18}$; $p_C = p(\text{no avería}) = \dfrac{15}{18}$.

7.4.2. $L = n \ln 2 - 2n \ln \theta + \Sigma \ln (\theta - x_i)$; $\partial L/\partial\theta = 0 = -\dfrac{2n}{\theta} + \Sigma \dfrac{1}{\theta - x_i}$. La solución debe encontrarse numéricamente.

7.4.3. $L = n \ln \theta + (\theta - 1) \Sigma \ln x_i$. Por tanto, $\Sigma \ln x_i$ es suficiente. $\partial L/\partial\theta = 0 = n\theta^{-1} +$ $+ \Sigma \ln x_i$; $\hat{\theta} = -n/\Sigma \ln x_i$; Var $(\hat{\theta}) = n/(\Sigma \ln x_i)^2$.

7.4.4. $L = -n \ln \theta - \dfrac{1}{2\theta^2} \Sigma (x - \theta)^2$; como $\Sigma(x - \theta)^2 = x^2 + \theta^2 - 2\theta \Sigma x$. El estadístico suficiente tiene dimensión dos, es Σx^2 y Σx. El estimador *MV* es la solución de $\dfrac{\partial L}{\partial \theta} = 0$; $\hat{\theta}_{MV} = \dfrac{\bar{x}}{2} \pm \sqrt{\dfrac{5}{4}\bar{x}^2 + s^2}$.

7.4.5. $L = -n\,\ell n\,\sigma - \dfrac{1}{2\sigma^2} (\ell n\, x_i - \mu)^2$; $\dfrac{\partial L}{\partial \mu} = \dfrac{\Sigma(\ell n\, x_i - \mu)}{\sigma^2} = 0 \Rightarrow \hat{\mu} = \dfrac{\Sigma\, \ell n\, x_i}{n}$

$\dfrac{\partial L}{\partial \sigma^2} = \dfrac{-n}{2\sigma^2} + \dfrac{(\ell n\, x_i - \mu)^2}{2\sigma^4} = 0 \Rightarrow \hat{\sigma}^2 = \Sigma(\ell n\, x_i - \hat{\mu})^2/n$

7.4.6. $l(\theta; x) = 1/(b - a)^n$ para máx $\{x_i\} < b$ y mín $\{x_i\} > a$, $l(\theta; x)$ será máximo para $(b - a)^n$ mínimo, esto es para $\hat{b} =$ máx $\{x_i\}$ y $\hat{a} =$ mín $\{x_i\}$.

7.4.7. $L = n \ln \theta + n\theta \ln x_0 - (\theta + 1) \Sigma \ln x_i$; $\partial L/\partial\theta = 0$; $\hat{\theta} = n/\Sigma \ln (x_i/x_0)$.

7.4.8. $L = n_1 \ln (\theta + 0{,}05) + n_2 \ln (0{,}9 - 2\theta) + n_3 \ln (0{,}05 + \theta)$; $\dfrac{\partial L}{\partial \theta} = 0$

conduce a $\hat{\theta} = 0{,}0072$.

7.4.9. $L = n \ln p + (\Sigma k_i - n) \ln (1 - p)$; $\hat{p} = \dfrac{n}{\Sigma k_i} = \dfrac{1}{k}$; $\dfrac{\partial^2 L}{\partial p^2} = -\dfrac{n}{p^2} +$

$+ (\Sigma k_i - n)/(1-p)^2$, Var $(\hat{\theta}) = \left(-\dfrac{\partial^2 L}{\partial p^2} \right)^{-1}_{\hat{p}} = \left[\dfrac{n}{\hat{p}^2} - \left(\dfrac{n}{\hat{p}} - n \right) / (1-\hat{p})^2 \right]^{-1}$

7.4.10. $\hat{\theta} = -n/\Sigma \ln (1 - x_i)$.

7.4.11. Si $f(x) = \lambda e^{-x\lambda}$; $p(x > 85) = e^{-\lambda 85}$; $l(\lambda) = \prod_{i=1}^{6} \lambda e^{-x_i \lambda} . (e^{-\lambda 85})^4$

$L(\lambda) = 6 \ln \lambda - \lambda \Sigma x_i + 4(-\lambda \cdot 85)$; $\hat{\mu} = \dfrac{1}{\hat{\lambda}} = \bar{x} + \dfrac{4}{6} 85 = 106{,}67$.

7.4.12. a) Sea $q_1 = 1 - p_1$; $q_2 = 1 - p_2$; $p(0) = q_1^3 q_2$; $p(1) = 3p_1 q_1^2 q_2 + q_1^3 p_2$
$p(2) = 3p_1^2 q_1 q_2 + 3p_1 p_2 q_1^2$; $p(3) = p_1^3 q_2 + 3p_1^2 q_1 p_2$; $p(4) = p_1^3 p_2$.
b) $E[x] = \Sigma x_i p_i = 3p_1 + p_2$; Var $(x) = 3p_1 q_1 + p_2 q_2$.
c) $p_1 = 0{,}05$; $p_2 = 0{,}09$.

7.4.13. $E[a\Sigma(x_i - \bar{x})^2] = a\sigma^2(n-1)$; Var $[a\Sigma(x_i - \bar{x})^2] = 2a^2\sigma^4(n-1)$;
$ECM = [\sigma^2 - a\sigma^2(n-1)]^2 + 2a^2\sigma^4(n-1)$. $\partial ECM/\partial a = 0$ implica $a = (n+1)^{-1}$.

7.4.14. $E(s^2) \to \sigma^2$ y Var $(s^2) \to 0$.

7.4.15. $\bar{x}/s$.

7.4.16. Si $f(x) = \alpha_1 N(\mu_1, \sigma_1) + \alpha_2 N(\mu_2, \sigma_2)$; $E(y) = \alpha_1\mu_1 + \alpha_2\mu_2$;
Var $(y) = \alpha_1\sigma_1^2 + \alpha_2\sigma_2^2 + (\mu_1 - \mu_2)^2 \alpha_1\alpha_2$.

7.4.17. a) $(s^2)^k$.
b) $E[(s^2)^k] = \dfrac{\sigma^{2k}}{n^k} E\left[\left(\dfrac{ns^2}{\sigma^2} \right)^k \right] = \dfrac{\sigma^{2k}}{n^k} E[(\chi^2)^k] = \sigma^{2k} \dfrac{h(n)}{n^k}$
c) El estimador $\dfrac{n^k}{h(n)} s^{2k}$ es centrado.

Capítulo 8

Ejercicios 8

8.1. La media es $(23 + 27 + 31 + 23 + 29 + 23 + 24 + 21 + 18 + 28)/10 = 24{,}7$ y la desviación típica $\hat{s} = 3{,}97$. El intervalo de confianza al 95% suponiendo normalidad como el valor de función de la distribución de la t con 9 grados de libertad es 2,26, el intervalo es $24{,}7 \pm 2{,}26 \cdot 3{,}97/\sqrt{10} = 24{,}7 \pm 2{,}84 = (21{,}9, 27{,}5)$. Concluimos que la edad media de los descubrimientos matemáticos está 21,9 años y 27,5 años con el 95% de confianza.

8.2. La media de estos trece datos es 34,31 y la desviación típica es 6,51. Suponiendo normalidad, el intervalo de confianza para la media es $34,31 \pm 2,18 \cdot \cdot 6,51\sqrt{13} = 34,31 \pm 3,94 = (30,37, 38,25)$.

8.3. Calcularemos el intervalo para la diferencia suponiendo que las poblaciones son normales con la misma varianza. Para calcular el intervalo de confianza para la diferencia entre edades medias estimamos la varianza

$$\hat{s}_T = \sqrt{\frac{9 \cdot 3,97^2 + 12 \cdot 6,51^2}{21}} = 5,57 \; 34,31 - 24,7 \pm 2,08 \cdot 5,57\sqrt{\frac{1}{10} + \frac{1}{13}} =$$

$$= 9,61 \pm 2,08 \cdot 5,57 \cdot 0,42 = 9,61 \pm 4,87 = (4,74; 14,48).$$

8.4. Supongamos que las varianzas son distintas. Entonces el intervalo de confianza para la diferencia entre edades medias es

$$34,31 - 24,7 \pm t \sqrt{\frac{3,97^2}{10} + \frac{6,51^2}{13}} = 9,61 \pm t2,2.$$

El valor de t se obtiene con $10 + 13 - 2 - \Delta$ grados de libertad donde $\Delta =$

$$= \frac{(9 \cdot 1,3^2 - 12 \cdot 1,8^2)^2}{9 \cdot 1,3^4 + 12 \cdot 1,8^4} = 3,69.$$

Por tanto la t tiene 17 grados de libertad y el intervalo es $9,61 \pm 2,11 \cdot 2,2 =$ (4,93; 14,29). Como vemos el intervalo es muy parecido al caso anterior.

8.5. El intervalo es $3,4 \pm 1,96 \cdot 1, 2/\sqrt{40} = (3,03; 3,77)$.

8.6. El intervalo es $2340 \pm 2,20 \cdot 815/\sqrt{12} = (2857; 1822)$.

8.7. Si la amplitud es 500 $L = 250$ y al 95% el valor de la normal es 1,96 y tendremos $n = 1,96^2 \cdot 815^2/250^2 = 40,8$. Luego al menos 41 estaciones.

8.8. $p_1 = 0,76$ y $p_2 = 112/140 = 0,80$. El intervalo es $p_1 - p_2 \in (0,76 - 0,80) \pm$

$$\pm 1,96 \sqrt{\frac{0,76 \cdot 0,24}{100} + \frac{0,8 \cdot 0,2}{140}} = -0,04 \pm 0,1068.$$

8.9. $\mu \in \bar{x} \pm 2,57 \dfrac{\sigma}{\sqrt{n}}$; $L = 2 \cdot 2,57 \dfrac{\sigma}{\sqrt{n}}$; $n = (2 \cdot 2,57 \cdot \sigma)^2/L_2$

8.10. Como σ_1 y σ_2 son conocidos $\bar{x}_1 - \bar{x}_2$ es $N\left(\mu_1 - \mu_2; \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}\right)$, luego:

$\mu_1 - \mu_2 \in (170,2 - 176,7 \pm 1,96\sqrt{225^2/10 + 256^2/12})$.

8.11. $\bar{x} = 1,5$; $\bar{y} = 2$; $s_1^2 = 1,11$; $s_2^2 = 36/11$; $\hat{s}_1^2/\hat{s}_2^2 = 1,11$; $\sigma_1^2/\sigma_2^2 < 3,48$.

8.12. $y = \Sigma x_i$; $E[y] = 25 \cdot 4 = 100$ $\text{Var}(y) = 25 \cdot 0{,}5^2/12 = 0{,}52$; $y \sim$ Normal, intervalo $100 \pm 1{,}96 \cdot 0{,}722$.

8.13. a) $(\hat{p} - p)/\sqrt{\hat{p}\hat{q}/n}$ es $N(0, 1)$ aproximadamente;
$A = 0{,}55 \pm 1{,}96\sqrt{0{,}55 \cdot 0{,}45/100}$; $B = 0{,}45 \pm 1{,}96\sqrt{0{,}55 \cdot 0{,}45/10}$.
b) $1{,}96 \cdot 0{,}497/\sqrt{n} < 0{,}5$; $n \geqslant 380$.

8.14. $\alpha = 0{,}05$; $F_{9,9} = 3{,}18$; $\sigma_2^2/\sigma_1^2 \leqslant 3{,}18 \cdot (0{,}6065/0{,}5919) = 3{,}26$.

8.15. a) Con $\alpha = 0{,}01$, $F_{120,\,120;\,0{,}005} = 1{,}6055$:
$(25/30)^2(1/1,\ 6055) \leqslant \sigma_1^2/\sigma_2^2 \leqslant (25/30)^2\ 1{,}6055$; intervalo $(0{,}4325;\ 1{,}1149)$.
b) Supuesto $\sigma_1^2 = \sigma_2^2$, $\hat{s}_T = 2.761$; intervalo $107 \pm 916{,}2$.

8.16. a) $\hat{s}_T = 229{,}48$; $t_{18;\,0{,}025} = 2{,}101$; $\alpha = 0{,}05$, suponiendo igualdad $(136{,}7 \pm \pm\ 217{,}107)$.
b) $g = 17{,}08 \simeq 17$; $(1362{,}5 - 1225{,}8 \pm 2{,}11 \cdot 103{,}33)$.

8.17. a) $L(\lambda) = -n\lambda + n\bar{x} \ln \lambda$; $\partial^2 L/\partial^2\lambda = -n\bar{x}/\lambda^2$; como $\hat{\lambda}_{MV} = \bar{x}$.

$$\text{Var}(\bar{x}) = \frac{\bar{x}}{n} \text{ luego } \frac{\bar{x} - \lambda}{\sqrt{\bar{x}}/\sqrt{n}} \text{ es } N(0, 1)\ ;\ \lambda = \bar{x} \pm z_{\alpha/2}\ \sqrt{\bar{x}}/\sqrt{n}.$$

b) $L(\lambda) = n \ln \lambda - \lambda n\bar{x}$; $\partial^2 L/\partial\lambda^2 = -n/\lambda^2$; $\hat{\lambda}_{MV} = \bar{x}^{-1}$;

$$\text{Var}(\hat{\lambda}_{MV}) = (\bar{x}^{-2}/n);\ \frac{\bar{x}^{-1} - \lambda}{\bar{x}^{-1}/\sqrt{n}} \text{ es } N(0, 1),\ \lambda = \bar{x}^{-1} \pm z_{\alpha/2}\bar{x}^{-1}/\sqrt{n}.$$

Capítulo 9

Ejercicios 9

9.1. $n_0 = \dfrac{1{,}8^2}{3^2} = 0{,}36$; $\mu_p = (25/25{,}36)\ 18 + (0{,}36/25{,}36)\ 15 = 17{,}95$
$\sigma_p = 1{,}80/\sqrt{25{,}36} = 0{,}357$. Intervalo $17{,}95 \pm (1{,}96)\ (0{,}357)$.

9.3. $1.700 = \alpha\ 1.800 + (1 - \alpha)\ 1.500$; $\alpha = 200/300 = 2/3 = 25/(25 + n_0)$; $n_0 = 12{,}5$; $\sigma_0 = 180/\sqrt{12{,}5} = 51$.

9.4. $P(p = 0{,}05|x) = k \cdot P(x|0{,}05)\ 0{,}8 = k \cdot 0{,}95^3 \cdot 0{,}8 = k \cdot 0{,}6859$; $P(p = 0{,}1|x) = k \cdot \cdot\ p(x|0{,}1) \cdot 0{,}2 = k \cdot 0{,}9^3 \cdot 0{,}2 = k \cdot 0{,}1458$, donde $k^{-1} = P(x)$. Entonces $k \cdot\ 0{,}6859 + + k \cdot\ 0{,}1458 = 1$; $k = 1{,}20$; $P_f(0{,}1) = 0{,}175$, $P_f(0{,}05) = 0{,}825$.

9.5. $f(\vartheta) = 10^{-1}$; $f(x|\vartheta) = k(x - \frac{1}{2}; x + \frac{1}{2})$; $f(x|\vartheta) = 0$ si $\vartheta \notin (x \pm \frac{1}{2})$ $f(\vartheta|x) = k\ 11 \leq 9$.

9.6. Tómese $A = n/\sigma^2$, $a = \bar{x}$, $B = 1/\sigma_0^2$, $b = \mu_0$, $x = \mu$ y aplíquese la igualdad que se demuestra desarrollando ambos miembros.

Capítulo 10

Ejercicios 10.1

10.1.1. a) Simple; b), c) y d) compuestas.

10.1.2. $d > 6{,}635$.

10.1.3. Como $p(x_1^2 \leq 5{,}023) = 0{,}975$, $p = 0{,}025$.

10.1.4. $\bar{x} \sim N(\lambda; \sqrt{\lambda/n})$, por tanto $d = \dfrac{\bar{x} - \lambda}{\sqrt{\lambda/n}} \sim N(0, 1)$. Con $\alpha = 0{,}05$

$$\left| \frac{\bar{x} - 5}{\sqrt{5/200}} \right| \leq 1{,}96.$$ Por tanto, rechazo si $|\bar{x} - 5| > 1{,}96\sqrt{5/200}$

10.1.5. $\bar{x} = \hat{p} \sim N(p; \sqrt{pq/n})$. Con $d = \dfrac{\hat{p} - 0{,}08}{\sqrt{0{,}08 \cdot 0{,}92/100}} \sim N(0, 1)$,

por tanto: $|\hat{p} - 0{,}08| > 1{,}96\sqrt{0{,}08 \cdot 0{,}92/100}$ región de rechazo para $\alpha = 0{,}05$.

10.1.6. $\bar{x} \sim N(1/\lambda; 1/\lambda\sqrt{n})$. Hipótesis $1/\lambda = 300$ horas.
$d = (250 - 300)(\sqrt{100}/300) = -1{,}67 \cdot p(|d| \geq 1{,}67) = 0{,}094$, que es el p crítico.

Ejercicios 10.2

10.2.1. $\bar{x}_1 = 130$, $\bar{x}_2 = 120$, $t_{12}(0{,}05) = 2{,}179$ y $d = 0{,}51$. La región de rechazo es: $|d| > 2{,}179$ como $d = 0{,}51$ se acepta.

10.2.2. $s^2 = 260/15$. $H_0 : \sigma^2 = 9$, si H_0 cierto $15s^2/9$ es χ^2 (14).
a) $\chi^2(14; 0{,}05) = 23{,}68$. Como $260/9 = 28{,}89 > 23{,}68$ se rechaza H_0.
b) $p = 0{,}01$.

10.2.3. $x \sim N(1.800, \sigma)$; $t = 4$, $t(9; 0{,}01) = 3{,}25$ luego $p < 0{,}01$.

10.2.4. La región de rechazo será $\bar{x} > \mu_0 + \sigma/\sqrt{n}\ z_\alpha$: para que $\alpha = \beta$
$\mu_0 + (\sigma\sqrt{n})z_\alpha = \mu_0 + \delta + z_{1-\beta}\sigma/\sqrt{n}$, es decir $\sqrt{n} = \sigma(z_\alpha - z_{1-\beta})/\delta$.

10.2.5. $\alpha = P(x \geq 4|\mu = 1) = 1 - \sum_0^3 e^{-1}/x! = 1 - 8/3e$; $\beta = P(x < 4|\mu = 2) = 19/3e^2$.

10.2.6. $\alpha = 1/3$; $\beta = 3/5$; Potencia = 2/5.

10.2.7. a) $\hat{s}_T = 115{,}75$, $t = 3{,}16$ las medias son distintas aunque las varianzas teóricas son iguales $F = (119{,}47/111{,}91)^2 = 1{,}15$.
b) Michelson $t = (896 - 792{,}5\sqrt{15}/111{,}91) = 3{,}58$, tiene sesgo; Newcomb $t =$ $= + 0{,}061 \cdot 0{,}368 = 0{,}894$.

10.2.8. a) $t_4 = (198{,}8 - 190)\sqrt{5}/10{,}71 = 1{,}84$; $t_4(0{,}05) = 2{,}78$ se acepta H_0.
b) $4 \cdot 10{,}71^2/100 = 4{,}58$, se acepta H_0.
c) Potencia $(\mu) = P[(190 - \mu)\sqrt{5}/10 - 1{,}96 < z < (190 - \mu)\sqrt{5}/10 + 1{,}96]$, donde z es $N(0, 1)$. Dando valores μ se obtiene $P(\mu)$.

10.2.9. $F_{9,\,11} = (2100/1900)^2 = 1{,}22$, se acepta igualdad de varianza; H_0: $\mu_A - \mu_B \leq 2.000$; H_1: $\mu_A - \mu_B > 2.000$; $\hat{s}_T = 1.992{,}49$, $t = 3{,}52$. Como $t_{20}(0{,}01) = 2{,}52$ se rechaza H_0 con $p < 0{,}01$ y compraremos A.

10.2.10. a) Se observa $z = (13{,}8 - 15)/(2/5) = -3 < -2{,}326$ y se rechaza H_0, la región de rechazo es $\bar{x} < 14{,}0696$; b) potencia $(14) = P\left(z < \dfrac{14{,}0696 - 14}{0{,}4}\right) = 0{,}569$. La potencia para cualquier valor de μ es $P[z < (14{,}0696 - \mu)/0{,}4]$.

10.2.11. $\bar{x}_A = 26{,}25$; $\bar{x}_B = 27{,}75$; $s_T^2 = 7{,}25$; $|t_6| = 0{,}788$, como $t(6; 0{,}05) = 2{,}447$ se acpeta H_0.

10.2.12. $|z| = |0{,}51 - 0{,}55|/\sqrt{0{,}51 \cdot 0{,}49/1.000} = 2{,}53$. Como con $\alpha = 0{,}01$ rechazo es $|z| > 2{,}576$ no puede rechazarse $p = 0{,}55$ con $\alpha = 0{,}01$, aunque sí con $\alpha = 0{,}05$.

Ejercicios 10.3

10.3.1. $\lambda = [(1/\bar{x})^n e^{-n}]/[e^{-n\bar{x}}] = e^{n\bar{x}} \cdot e^{-n}/\bar{x}^n$ y $\lambda > k$ implica $\bar{x} < k$. Para determinar k, $2 \ln \lambda = 2n\bar{x} - n - n \ln \bar{x}$) es $\chi^2(1)$. Rechazo es: $\bar{x} - \ln \bar{x} > 1{,}192$.

10.3.2. El estimador MV de ρ es $r = s_{xy}/s_x s_y$. Entonces $\ell\,(\hat{\Omega}_0) = \dfrac{1}{(2\pi s_1 s_2)^n}\, e^{-n}$

$$\ell\,(\hat{\Omega}_0) = \frac{1}{(2\pi s_1 s_2)^n}\,(1 - r^2)^{-n/2} e^{-n} \text{ y } \lambda = (1 - r^2)^{-n/2};\ \lambda > k \Rightarrow |r| > C.$$

Como $2 \ln \lambda \sim \chi^2\,(1)$, rechazo: $-n \ln (1 - r^2) > 3{,}84$; $(1 - r^2) < 0{,}83$.

10.3.4. $\lambda = (\hat{p}/p_0)^r (\hat{q}/q_0)^{n-r}$ con $\hat{p} = r/n$; $\hat{q} = 1 - \hat{p}$. Por tanto:

$$\lambda = \left(\frac{r}{np_0}\right)^r \left(\frac{n - r}{nq_0}\right)^{n-r},\ \lambda > k \Rightarrow a > r > b.$$

10.3.5. $\ell\,(\hat{\Omega}_0) = \begin{cases} 0 & \text{si } x_{\text{máx}} > 2 \\ \dfrac{1}{4}\,x_i \text{ si } x_{\text{máx}} < 2 \end{cases}$;

$$\ell\,(\hat{\Omega}) = \begin{cases} \dfrac{1}{4}x_i & \text{si } x_{\text{máx}} > 2 \\ \dfrac{2^n \pi x_i}{(x_{\text{máx}})^{2/n}} \text{ si } x_{\text{máx}} < 2 \end{cases}$$

Por tanto, la zona de rechazo es: si $x_{\text{máx}} > 2$, $\lambda = 0$ (rechazar siempre);

si $x_{\text{máx}} < 2$, $\lambda = \dfrac{1}{4}\,\dfrac{(x_{\text{máx}})^{2/n}}{2/n} \leq k$.

k se determina con $2 \ln \sim \chi^2\,(1) \Rightarrow$ Rechazo si $0{,}2 \ln x_{\text{máx}} \geq 12{,}79$.

Capítulo 11

Ejercicios 11

11.1. EM(Perforar) = 0,5(–20) + 0,3(10) + 0,2(100) + 13 > 0, luego conviene perforar.

11.2. $BEIP = 0{,}4(0) + 0{,}3(10) + 0{,}2(100) = 23$, luego $VEIP = 23 - 13 = 10$.

11.3. $P(\text{no}) = P(\text{no}|N)P(N) + P(\text{no}|P + M)P(P + M) = 0{,}5$; $P(s) = 0{,}5$. Si prueba es no: $P(N|n) = P(n|N)P(N)/P(n) = 0{,}8$; $P(P|n) = P(n|P)P(P)/P(n) = 0{,}12$; $P(M|n) = P(n|M)P(M)/P(n) = 0{,}08$; luego $BE(\text{Perforar}|n) = -6{,}8$ y es mejor no perforar. Si prueba es sí, $P(N|s) = 0{,}2$; $P(P|s) = 0{,}48$; $P(M|s) = 0{,}32$; luego $BE(\text{Perforar}|s) = 32{,}8$. Por tanto $BE(\text{Prueba}) = 0{,}5(0) + 0{,}5(32{,}8) = 16{,}4$. $VEIM = 16{,}4 - 13 = 3{,}4$ es la cantidad máxima a pagar.

11.4. $u(-20) = -267{,}99$; $u(0) = 0$; $u(10) = 4{,}32$; $u(100) = 4{,}99$ y es mejor no perforar, ya que $u(\text{perforar}) = -131{,}7$.

11.5. $u(15) = 34{,}8$; $u(35) = -121{,}15$; $u(70) = -10.956$; $u(30) = -109$; $u(4) = -535{,}9$; $u(50) = -1.474{,}1$.

$u(A) = -3.482$; 88; $u(B) = -782$, y el camino B es preferido.

11.6. Si $r = -d \ln u'(x)/dx$; $-\ln u'(x) = k + rx$; $u'(x) = ke^{-rk}$. $u(x) = -1/rk\, e^{-rk} + C$, si $u(0) = 0 \Rightarrow C = k/r$, luego $u(x) = k(1 - e^{-rk})/r$ con k arbitraria.

11.7. $u'(x) = (x + a)^{-1}$; $u''(x) = -(x + a)^{-2}$; $r(x) = (x + a)^{-1}$.

Capítulo 12

Ejercicios 12.1

12.1.1. Se acepta la normalidad.

12.1.2. Distribución de Poisson. Se acepta.

12.1.3. Exponencial. Se acepta.

12.1.4. Con Kolmogorov-Smirnov $D_n = 0{,}3715$, equivale a $\alpha \simeq 0{,}1$.

12.1.5. Con $\chi^2 = 8{,}5$ con 5 grados de libertad. Se acepta equiprobabilidad.

12.1.6. Normal. Se acepta.

12.1.7. Con Kolmogorov-Smirnov $D_n = 0{,}17$; con $\alpha = 0{,}1$, $D_{\text{crítico}} = 0{,}338$. Se acepta normalidad.

12.1.8. Tomando logaritmos se acepta la hipótesis de normalidad.

12.1.9. $\overline{x} = 41{,}57$ ptas., $\hat{s} = 11{,}84$. La distribución es asimétrica. Tomando logaritmos, si $y = \ln x$, $\overline{y} = 3{,}69$, $\hat{s}_y = 0{,}278$. Con los mismos intervalos transformados $X^2 = 6{,}43$. Como $\chi^2(7;\, 0{,}95) = 14{,}1$ se acepta que los datos en logaritmos son normales.

12.1.10. $\overline{x} = 171{,}98$ cm, $s = 7{,}15$ cm, $X^2 = 12{,}5$ como $\chi^2(6;0{,}05) = 12{,}6$ acepta normalidad. (b) El intervalo del 95% para la media es $(171{,}98 \pm 0{,}625)$. (c) El intervalo no incluye 172,9; luego en esa junta están por debajo de la media.

12.1.11. $\overline{x} = 107{,}73$ minutos, $s = 18$. Son asimétricos. Transformando con $y = \ln x$, $\overline{y} = 4{,}66$, $\hat{s}_y = 0{,}16$. $X^2 = 2{,}336$, se acepta la normalidad en logaritmos.

12.1.12. $\overline{x} = 3.244{,}7$ m; $s = 1.351{,}1$ m. Como son asimétricos, con $y = \ln x$, $\overline{y} = 7{,}98$, $s = 0{,}45$. $X^2 = 12{,}83$ como $\chi^2(8;\, 0{,}05) = 15{,}5$, se acepta la normalidad en logaritmos.

12.1.13. Poisson con $\lambda = 0{,}69 \cdot x^2 = \dfrac{(223-217)^2}{217} + \dfrac{(142-149)^2}{149} + \ldots + \dfrac{(4-2)^2}{2} = 3{,}55$. Se acepta Poisson.

12.1.14. Poisson con $\lambda = 0{,}61 \cdot x^2 = \dfrac{(109-108{,}7)^2}{108{,}7} + \dfrac{(65-66{,}3)^2}{66{,}3} + \ldots + \dfrac{(1-0{,}6)^2}{0{,}6} = 0{,}748$. Se acepta Poisson.

Ejercicios 12.2

12.2.1. De $\Sigma(x_i - \bar{x})^2 = \Sigma(x_i - \mu)^2 - (\bar{x} - \mu)^2$ resulta $(n-1)E(\hat{s}^2) = n\sigma^2 - n\,\mathrm{Var}(\bar{x})$; sustituyendo la expresión del texto (3.2) para $\mathrm{Var}(\bar{x})$, $E(\hat{s}^2) = (n/n-1)\sigma^2 - \sigma^2[1/(n-1) + 2\rho/n] = \sigma^2(1 - 2\rho/n)$.

12.2.2. Como $\mathrm{Var}(\bar{x}) \geq 0 \Rightarrow 1 - 2\rho/n > 0 \Rightarrow \rho < n/2$, para $n = 1$, $\rho < 0{,}5$.

12.2.3. Ordenando los datos la mediana es 16 y la secuencia de signos es – – – – – + + + + + + – . Hay 3 rachas, de la tabla 12 concluimos que el test no es concluyente, dado el pequeño tamaño muestral. Comparar este resultado con el ejercicio 6.10.

12.2.4. $r_1 = 0{,}59$; $r_2 = 0{,}41$; $r_3 = 0{,}275$.

Ejercicios 12.3

12.3.1. $\dfrac{(42-53{,}5)^2}{53{,}5} + \dfrac{(58-46{,}5)^2}{46{,}5} + \dfrac{(65-53{,}5)^2}{53{,}5} + \dfrac{(35-46{,}5)^2}{46{,}5} = 10{,}63$ Como $\chi^2(1;\ 0{,}05) = 7{,}88$ admitimos que hay diferencias entre los sexos.

12.3.2. La χ^2 resulta es:
$\dfrac{(15-7{,}82)^2}{7{,}82} + \dfrac{(18-18{,}39)^2}{18{,}39} + \dfrac{(7-13{,}79)^2}{13{,}79} + \ldots + \dfrac{(23-16{,}21)^2}{16{,}21} = 18{,}43$. Hay dos grados de libertad. Como $\chi^2(0{,}05;\ 2) = 5{,}99$ rechazamos la hipótesis de homogeneidad. Las calificaciones son distintas en ambas asignaturas.

12.3.3. $\chi^2 = 0{,}01 + 1{,}67 + 11{,}52 + 0{,}06 + 0{,}48 + 1{,}66 + 0{,}11 + 0{,}10 + 1{,}96 = 17{,}58$ con 4 grados de libertad. $\chi^2(4;\ 0{,}05) = 9{,}49$; $\chi^2(4;\ 0{,}005) = 14{,}9$. Luego hay claramente diferencias significativas.

12.3.4. $\chi^2 = \dfrac{8{,}5^2}{31{,}5} + \dfrac{(-3{,}4)^2}{39{,}4} + \dfrac{(-5{,}1)^2}{7{,}1} + \dfrac{8{,}5^2}{8{,}5} + \dfrac{3{,}4^2}{10{,}6} + \dfrac{5{,}1^2}{1{,}9} = 29{,}37$. Como $\chi^2(2;\ 0{,}05) = 5{,}99$, rechazaremos la independencia, ambas medidas están relacionadas.

Capítulo 13

Ejercicios 13.1

13.1.1. a) Estimación inicial $\hat{\sigma} = 215{,}07/2{,}326 = 92{,}46$; gráfico de la media $119 \pm \pm 124{,}05$; gráfico del rango $(0;\ 454{,}9)$. Muestra 8 fuera de control. Al elimi-

narla, resulta $\bar{\bar{x}} = 107{,}87$; $\bar{R} = 206{,}50$ y $\hat{\sigma} = 88{,}8$. Los puntos están dentro de los nuevos gráficos.

b) Capacidad = $6 \cdot 88{,}8$.

13.1.2. a) Media: $431 \pm 33{,}15$; rangos: (0; 103,8).

b) $\hat{\sigma} = 22{,}10$.

c) $IC = 60/(6 \cdot 22{,}10) = 0{,}45$, proceso no apto; $P(400 < x < 460) = P[(400 - 431)/22{,}1 < z < (460 - 431)/22{,}1] = 0{,}8241$, donde $z \sim N(0, 1)$, luego un 17,6% defectuosos por estar fuera de 430 ± 30.

13.1.3. Los límites son $\mu \pm 3\sigma/\sqrt{4} = \mu \pm 1{,}5\sigma$. Si $\mu_1 = \mu \pm 0{,}5\ \sigma$, la prob. de un punto fuera de límites es $P(\bar{x} > \mu + 1{,}5\sigma/\bar{x} \sim N[\mu + 0{,}5\sigma;\ \sigma/2]) = P(z > 2) = 0{,}0228$ (donde $z \sim N[0, 1]$). Harán falta $1/0{,}0228 = 43{,}86$ observaciones en promedio para detectarlo, luego 43,9 h.

13.1.4. Sea $\beta(k)$ la probabilidad de que esté en límites cuando $\mu = \mu_0 + k\sigma$;
$\beta(k) = p(\mu_0 - 3\sigma/\sqrt{9} < \bar{x} < \mu_0 + 3\sigma/\sqrt{9}) =$
$= p[(\mu_0 - \sigma - \mu_0 - k\sigma)/\sigma < (\bar{x} - \mu_0 - k\sigma)/\sigma < (\mu_0 + \sigma - \mu_0 - k\sigma)/\sigma] =$
$= p(-1 - k < z < 1 - k)$, donde $z \sim N(0, 1)$.

13.1.5. $\hat{p} = 0{,}12$. Límites son $23{,}1 \pm 3 \cdot 4{,}52 = (9{,}54;\ 36{,}66)$. Observaciones 10, 13 y 14 fuera de control. Al eliminarlas $\hat{p} = 0{,}10$, límites $19{,}88 \pm 3 \cdot (19{,}88 \cdot 0{,}9)^{0{,}5} = 19{,}88 \pm 3 \cdot 4{,}23 = (7{,}19;\ 32{,}57)$. El punto 12 está fuera de control. Al eliminarlo $\hat{p} = 0{,}095$ y todos están dentro de los límites.

13.1.6. $\hat{c} = 3{,}6$; por tanto, el gráfico tendrá límites $3{,}6 \pm 3\sqrt{3{,}6} = (0;\ 9{,}29)$.

13.1.7. Los límites de control para las 5 primeras sería $(1{,}44 \pm 3\sqrt{1{,}44})$ pero las quince siguientes $1{,}44 \pm 33\sqrt{1{,}44/3}$.

13.1.8. a) $6{,}2 \pm 3\sqrt{6{,}2}$.

b) $49{,}6 \pm 3\sqrt{49{,}6}$.

13.1.9. Si x es un dígito, $E[x] = 4{,}5$; $\text{Var}(x) = (1^2 + 2^2 + ... + 9^2)/10 - 4{,}5^2 = 8{,}25$. Si $y = \sum_{1}^{100} x$; $E[y] = 450$; $\sigma(y) = \sqrt{825}$. Por tanto, límites $450 \pm 3\sqrt{825}$.

Ejercicios 13.2

13.2.1. Tomar 125 elementos, si hay 5 o menos defectuosos aceptar, si 6 o más rechazar (véase tabla 13.7).

13.2.2. a) Mil-Std: AOQ = 0,01; en tablas $n = 200$, rechazo si $r \geq 6$.
b) $D - R$; en tablas $n = 290$; $r > 7$; AOQL = 1,4%.
c) Comparación:
Mil-Std: $\alpha = P(r \geq 6/p = 0{,}01) = P(r \geq 6/\lambda = 0{,}01 \cdot 200) = P[(r - 2)/\sqrt{2} \geq \geq 2{,}47] = 0{,}01$. $\beta = P(r < 6/\lambda = 8) = P(z < -0{,}88) = 1 - 0{,}8106 \simeq 0{,}19$.
$D - R$: $\alpha = P(r > 7/\lambda = 2{,}9) = P(z > 2{,}70) = 0{,}004$; $\beta = P(z < -1{,}20) = = 1 - 0{,}8849 \simeq 0{,}12$.

13.2.3. M.S.: $n = 100$. Rechazar si $c > 5$; DR: $n = 260$. Rechazar si $c > 6$. Calculando las curvas *OC* para ambos planes es mejor DR para el comprador.

13.2.4. $n = \sigma^2(z_{1-\alpha} - z_\beta)^2/(\mu_1 - \mu_0)^2$; $n = 9$, $a = 200 + 1{,}64 \cdot 10/\sqrt{9}$.

13.2.5. Utilizando la aproximación de Poisson, con $p = 0{,}003$; $\lambda = 0{,}3$, $OC(p) = = 0{,}9964$; AOQ = $2{,}99 \cdot 10^{-3}$; $p = 0{,}1$; $\lambda = 1$, $p(A/\lambda) = 0{,}9197$; AOQ = $9{,}2 \cdot \cdot 10^{-3}$; $p = 0{,}02$; $\lambda = 2$; $P(A/\lambda) = 0{,}6767$; AOQ = $13{,}5 \cdot 10^{-3}$; $p = 0{,}03$; $\lambda = 3$; $P(A/\lambda) = 0{,}423$; AOQ = $12{,}7 \cdot 10^{-3}$, etc.

13.2.6. a) $P(A/p) = (1 - p)^5$.
b) $\text{AOQ}(p) = p(1 - p)^5$.
c) 0,067 para $\hat{p} = 0{,}17$.
d) $(1 - p_1)^5 = 0{,}95$, $p_1 = 0{,}0102$.
e) $(1 - p_2)^5 = 0{,}1$; $p_2 = 0{,}369$.

13.2.7. a) Aproximando con Poisson, $\lambda = 1$, p(aceptar) = $0{,}736 + 0{,}184 \cdot 0{,}736 + + 0{,}061 \cdot 0{,}368 = 0{,}894$.
b) $E[n] = 50 \cdot 0{,}736 + 100 \cdot 0{,}135 + 100 \cdot (0{,}049 + 0{,}022 + 0{,}039) + 50 \cdot 0{,}019 = = 62{,}25$.

13.2.8. a) $P(A) = (1 - p)^5 + 5p(1 - p)^9 = (1 - p)^5[1 + 5p(1 - p)^4]$.
b) $\text{AOQ}(p) = p(1 - p)^5[1 + 5p(1 - p)^4]$.
c) AOQL $\simeq 0{,}095$ para $p = 0{,}18$.
d) 0,04.
e) 0,40.

Bibliografía

Los libros sin asterisco utilizan un nivel matemático similar o inferior al utilizado en este texto, mientras que los marcados con un asterisco requieren matemáticas más avanzadas.

a) Divulgación Estadística

Bartholomew, D. J. y Bassett, E. (1971): *Let's Look at the Figures.* Pelican, Penguin Books.

Borel, E. (1998): *Las Probabilidades y La Vida*. Yenny Promocion.

Huff, D. (1993): *How to Lie With Statistics.* W. W. Norton & Company.

Kapadia, R. y Andersson, G. (1987): *Statistics Explained: Basic Concepts and Methods.* Pearson Higher Education.

Kitaigorodski, A. (1976): *Lo inverosímil no es un hecho.* Editorial MIR.

Moroney, M. (1990): *Facts from Figures.* Pelican, Penguin Books.

Rao, C. (2004): *Estadística y verdad.* Edicions Universitat Barcelona.

Tanur, J. M. (2007): *La estadística : una guía de lo desconocido.* Alianza Editorial.

b) Historia de la estadística

Bernstein, P. L. (1998): *Against the Gods: The Remarkable Story of Risk.* John Wiley & Sons.

Box, J. F. (1978): *R.A. Fisher, The Life of a Scientist.* John Wiley & Sons.

David, F. N. (1998): *Games, Gods and Gambling: A History of Probability and Statistical Ideas.* Dover Publications, unabridged edición.

Gani, J. (1982): *The Making of Statisticians.* Springer.
Hald, A. (1990): *A History of Probability and Statistics and Their Applications before 1750.* Wiley-Interscience.
— (1998): *A History of Mathematical Statistics from 1750 to 1930* . Wiley-Interscience.
Kendall, M. G. (1977): *Studies in the history of statistics and probability: Volume II : a series of papers.* MacMillan.
Krüger, L., Daston, L. J., y Heidelberger, M., editores (1990a): *The Probabilistic Revolution, Volume 1: Ideas in History.* The MIT Press.
Krüger, L., Gigerenzer, G., y Morgan, M. S., editores (1990b): *The Probabilistic Revolution, Volume 2: Ideas in the Sciences.* The MIT Press.
Pearson, E. y Kendall, S. M., editores (1976): *Studies in the history of statistics and probability: A series of papers, Vol. I.* Hodder Arnold.
Pearson, E. S., *et al.* (1990): *Student: A Statistical Biography of William Sealy Gosset.* Oxford University Press.
Porter, T. M. (1988): *The Rise of Statistical Thinking, 1820-1900.* Princeton University Press.
Reid, C. (1982): *Neyman - From Life* (1ª edición). Springer.
Sánchez, C. y Valdés, C. (2003): *Kolmógorov, el zar de azar.* Nivola Libros y Ediciones, S.L.
Sánchez-Lafuente, J. (1975): *Historia de la estadística como ciencia en España 1500-1900.* Instituto Nacional de Estadística.
Stigler, S. M. (1990): *The History of Statistics: The Measurement of Uncertainty before 1900.* Belknap Press.
— (2002): *Statistics on the Table: The History of Statistical Concepts and Methods.* Harvard University Press.
Todhunter, I. (2007): *A history of the mathematical theory of probability: from the time of Pascal to that of Laplace (1865).* Kessinger Publishing, LLC.

c) Manuales de estadística general

Arnold, S. F. (1990): *Mathematical Statistics.* Prentice Hall.
*Bickel, P. J. y Doksum, K. A. (2006): *Mathematical Statistics, Updated Printing* (2ª edición). Prentice Hall.
Clarke, G. M. y Cooke, D. (2005): *A Basic Course in Statistics* (5ª edición). Hodder Arnold.
De La Horra, J. (2003): *Estadística Aplicada.* Ediciones Díaz de Santos.
DeGroot, M. H., *et al.* (1988): *Probabilidad y estadística.* Addison Wesley.
Dudewicz, E. J. y Mishra, S. N. (1988): *Modern Mathematical Statistics.* John Wiley & Sons.
Freedman, D., Pisani, R., y Purves, R. (2007): *Statistics* (4ª edición). W. W. Norton.
Garcia Pérez, A. *et al.* (1995). *Estadística I*, UNED
Graybill, F. A., Mood, A. M., y Poch, F. A. (1965): *Introducción a la teoría de la Estadística.* Aguilar.
Guttman, I. *et al.* (1982): *Introductory Engineering Statistics.*(3ª edición). John Wiley & Sons.
Harnett, D. L. y Murphy, J. L. (1987): *Introducción al análisis estadístico.* Addison Wesley Iberoamericana.

Hogg, R. V. y Ledolter, J. (1992): *Engineering Statistics.* (2ª edición) Prentice Hall.
Infante Gil, S. y Zárate de Lara, G. P. (1998): *Métodos estadísticos: Un enfoque interdisciplinario.* Trillas.
Larsen, R. J. y Marx, M. L. (2005): *An Introduction to Mathematical Statistics and Its Applications* (4ª edición). Prentice Hall.
Larsen, R. J., Marx, M. L., y Cooil, B. (1997): *Statistics for Applied Problem Solving and Decision Making.* South-Western College Pub.
Lindgren, B. (1993): *Statistical Theory.* (4ª edición). Chapman & Hall/CRC.
Maronna, R. A. (1995): *Probabilidad y Estadística Elementales para Estudiantes de Ciencias.* Editorial Exacta.
Newbold, P., Carlson, W. L., y Thorne, B. (2006): *Statistics for Business and Economics and Student CD* (6ª edición). Prentice Hall.
Peters, W. (1987): *Counting for Something: Statistical Principles and Personalities.* Springer-Verlag.
Rohatgi, V. K. (1976): *An Introduction to Probability Theory and Mathematical Statistics*. John Wiley & Sons.
Sachs, L. (1984): *Applied Statistics: A Handbook of Techniques.* (2ª edición). Springer.
Sarabia, J. M. (1993): *Curso práctico de Estadística.* Civitas.
Ugarte, M.D., Militino, A. F. y Arnholt, A. T. (2008): *Probability and Statistics with R*. CRC Press
Wackerly, D. D., Mendenhall, W., y Scheaffer, R. L. (2002): *Estadística matemática con aplicaciones.* Cengage Learning Editores.
Webster, A. L. (2005): *Estadística aplicada a los negocios y la Economía.* McGraw-Hill.
Wonnacott, T. H. y Wonnacott, R. J. (2004): *Introducción a la estadística.* Limusa.

d) Análisis de datos

Alegre, J. (1999): *Aplicaciones económicas de estadística descriptiva: ¿qué es y qué hacen los economistas con la estadística descriptiva?* Universitat de les Illes Balears. Servei de Publicacions.
Barnett, V. (1981): *Interpreting Multivariate Data*. John Wiley and Sons.
Cleveland, W. S. (1993): *Visualizing Data* (1ª edición). Hobart Press.
— (1994): *The Elements of Graphing Data.* (2ª edición). Hobart Press.
Chambers, J. M., Cleveland, W. S., y Tukey, P. A. (1983): *Graphical methods for data analysis.* Duxbury Press.
Ehrenberg, A. S. C. (1986): *A Primer in Data Reduction.* John Wiley & Sons.
Escuder, R. (1986): *Estadística económica y empresarial.* Editorial Tébar, S.L.
Fairley, W. B. y Mosteller, F. (1977): *Statistics and public policy.* Addison Wesley Longman Publishing Co.
*Flury, B. (1997): *A First Course in Multivariate Statistics.* Springer.
Hoaglin, D. C., Mosteller, F., y Tukey, J. W., editores (1985): *Exploring Data Tables, Trends, and Shapes.* John Wiley & Sons.
— (2000): *Understanding Robust and Exploratory Data Analysis.* Wiley-Interscience.
*Krzanowski, W. J. (2000): *Principles of Multivariate Analysis: A User's Perspective.* Oxford University Press.
Lebart, L., Morineau, A., y Fénelon, J. P. (1985): *Tratamiento estadístico de datos: métodos y programas.* Macombo.

Martín-Guzmán, P., y Martín Pliego, F. J. (1991): *Curso básico de estadística económica.* AC.
Martín Pliego, F. J. y García, M. (2004): *Introducción a la estadística económica y empresarial (teoría y práctica)*. Cengage Learning Editores.
Montiel Torres, A. M., Rius Díaz, F., y Barón López, F. J. (1996): *Elementos básicos de estadística económica y empresarial.* Prentice Hall.
Mosteller, F., Fienberg, S., y Rourke, R. E. K. (1983): *Beginning Statistics With Data Analysis.* Addison Wesley.
Mosteller, F. y Tukey, J. W. (1977): *Data Analysis and Regression: A Second Course in Statistics.* Addison Wesley.
Peña, D. (2002): *Análisis de datos multivariantes.* McGraw-Hill.
Tufte, E. R. (2001): *The Visual Display of Quantitative Information* (2ª edición). Graphics Press.
Tukey, J. W. (1977): *Exploratory Data Analysis.* Addison Wesley.
Velleman, P. F. y Hoaglin, D. C. (1981): *Applications, basics, and computing of exploratory data analysis.* Duxbury Press.

e) Cálculo de probabilidades

Allen, A. O. (1990): *Probability, Statistics, and Queuing Theory With Computer Science Applications* (2ª edición). Academic Press.
*Castillo, E. (1978): *Introducción a la estadística aplicada.* Publicado por el autor.
Cramér, H. (1968): *Elementos de la teoría de probabilidades y algunas de sus aplicaciones.* Aguilar.
De Finetti, B. (1974): *Theory of Probability. A Critical Introductory Treatment. Volume 1.* John Wiley and Sons.
Devroye, L. (1986): *Non-Uniform Random Variate Generation.* Springer.
*Feller, W. (1971): *An Introduction to Probability Theory and Its Applications, Volume 2.* John Wiley & Sons.
Gnedenko, B. V. (1998): *Theory of Probability* (6ª edición). CRC.
Johnson, N. L. y Kotz, S. (1972): *Distributions in Statistics 4 Vol.* John Wiley & Sons.
Kelly, D. G. (1993): *Introduction to Probability.* Prentice Hall.
*Kolmogorov, A. (1956): *Foundations of the Theory of Probability.* Chelsea Publishing Company.
Lindley, D. V. (1970): *Introduction to Probability and Statistics from a Bayesian Viewpoint. Part 1: Probability; Part 2: Inference.* Cambridge University Press.
*Loéve, M. (1976): *Teoría de la probabilidad.* Editorial Tecnos.
O'Hagan, A. (1988): *Probability Methods and Measurement.* Kluwer Academic Publishers.
Olkin, I., Gleser, L. J., y Cyrus, D. (1980): *Probability Models and Applications.* MacMillan.
Papoulis, A. y Pillai, S. (2002): *Probability, Random Variables and Stochastic Processes* (4ª edición). McGraw Hill.
Parzen, E. (1987): *Teoría moderna de probabilidades y sus aplicaciones.* Limusa.
Patel, J. K. y Read, C. B. (1996): *Handbook of the Normal Distribution* (2ª edición). CRC.
Quesada, V. y García Pérez, A. (1988): *Lecciones de cálculo de probabilidades.* Ediciones Díaz de Santos.

Rényi, A. (1966): *Calcul des probabilités.* Dunod.
Ross, S. (2005): *A First Course in Probability* (7ª edición). Prentice Hall.
*Springer, M. D. (1979): *The Algebra of Random Variables.* John Wiley & Sons.
Taylor, L. D. (1974): *Probability and mathematical statistics.* Harper & Row.
Trivedi, K. S. (2002): *Probability and Statistics with Reliability, Queing and Computer Science Applications* (2ª edición). Wiley-Interscience.
Vélez, R. (2004): *Cálculo de probabilidades 2.* Ediciones Académicas.

f) Técnicas de muestreo

Azorín, F. y Sánchez-Crespo, J. L. (1986): *Métodos y aplicaciones del muestreo.* Alianza.
Cochran, W. G. (1980): *Técnicas de muestreo.* Continental.
David, H., editor (1978): *Contributions to Survey Sampling and Applied Statistics: Papers in Honour of H.O.Hartley.* Academic Press.
Fernández, F. R. y Mayor, J. A. (1995): *Muestreo en poblaciones finitas: Curso básico.* EUB.
Kish, L. (1975): *Muestreo de encuestas.* Trillas.
Mirás, J. (1985): *Elementos de muestreo para poblaciones finitas.* Instituto Nacional de Estadística.
Sánchez-Crespo, J. L. (1984): *Curso intensivo de muestreo en poblaciones finitas.* Instituto Nacional de Estadística.
Sukhatme, P. V. (1984): *Sampling Theory of Surveys with Applications.* Iowa State University press.
Thompson, S. K. (2002): *Sampling.* John Wiley & Sons.

g) Inferencia paramétrica

*Barndorff-Nielsen, O. (1978): *Information and exponential families in statistical theory*. John Wiley & Sons.
Barndorff-Nielsen, O. E. y Cox, D. R. (1989): *Asymptotic Techniques for Use in Statistics.* Chapman and Hall.
Barnett, V. (1999): *Comparative Statistical Inference* (3ª edición). John Wiley & Sons.
Casella, G. y Berger, R. L. (2001): *Statistical Inference* (2ª edición). Duxbury Press.
*Cox, D. y Hinkley, D. (1979): *Theoretical Statistics.* Chapman & Hall/CRC.
Cuadras, C. M. (1996): *Métodos de análisis multivariante.* (3ª edición) EUB, Barcelona.
Edwards, A. W. F. (1992): *Likelihood (Expanded edition).* The Johns Hopkins University Press.
Hampel, F. R., *et al.* (2005): *Robust Statistics: The Approach Based on Influence Functions*. Wiley-Interscience.
*Huber, P. J. (1987): *Robust Statistical Procedures* (2ª edición). Society for Industrial Mathematics.
— (2003): *Robust Statistics*. Wiley-Interscience.
Johnson, R. A. y Wichern, D. W. (2007): *Applied Multivariate Statistical Analysis* (6ª edición). Prentice Hall.
*Lehmann, E. y Casella, G. (2003): *Theory of Point Estimation* (2ª edición). Springer.
*Lehmann, E. y Romano, J. P. (2008): *Testing Statistical Hypotheses* (3ª edición). Springer.

McLachlan, G. y Peel, D. (2000): *Finite Mixture Models*. Wiley-Interscience.
*Pitman, E. J. G. (1979): *Some basic theory for statistical inference*. Halsted Press.
*Prakasa Rao, B. L. S. (1987): *Asymptotic Theory of Statistical Inference*. John Wiley & Sons.
Rohatgi, V. K. (2003): *Statistical Inference.* Dover Publications.
Silvey, S. (1975): *Statistical Inference*. Chapman & Hall/CRC.
Stuart, A., *et al.* (2005): *Kendall's Advanced Theory of Statistics: 3-Volume Set* (6ª edición). Hodder Arnold.
Titterington, D. M. , Smith, A. F. M., y Makov, F. M. (1987): *Statistical Analysis of Finite Mixture Distributions.* John Wiley & Sons.
Zacks, S. (1971): *Theory of Statistical Inference.* John Wiley & Sons.

h) Diagnosis e inferencia no paramétrica

Barnett, V. y Lewis, T. (1994): *Outliers in Statistical Data* (3ª edición). John Wiley & Sons.
Breiman, L. (1973): *Statistics: with a view toward applications.* Houghton Mifflin.
Conover, W. J. (1999): *Practical Nonparametric Statistics* (3ª edición). John Wiley & Sons.
Efron, B. (1987): *The Jackknife, the Bootstrap, and Other Resampling Plans*. Society for Industrial Mathematics.
Efron, B. y Tibshirani, R. (1994): *An Introduction to the Bootstrap*. Chapman & Hall/CRC.
Everitt, B. S. (1992): *The Analysis of Contingency Tables* (2ª edición). Chapman & Hall/CRC.
Fienberg, S. (2007): *The Analysis of Cross-Classified Categorical Data* (2ª edición). Springer.
Hajek, J. (1969): *A Course in Nonparametric Statistics.* Holden-Day.
Lehmann, E. L. (2006): *Nonparametrics: Statistical Methods Based on Ranks.* Springer.
Mosteller, F. y Rourke, R. E. K. (1973): *Sturdy Statistics*. Addison Wesley Longman Publishing Co.
*Nadaraya, E. A. (1988): *Nonparametric Estimation of Probability Densities and Regression Curves.* Springer.
Noreen, E. W. (1989): *Computer-Intensive Methods for Testing Hypotheses: An Introduction.* Wiley-Interscience.
Read, T. R. y Cressie, N. A. (1988): *Goodness-of-Fit Statistics for Discrete Multivariate Data.* Springer.
Sarhan, A. E. y Greenberg, B., editores (1962): *Contributions To Order Statistics.* John Wiley and Sons.
Sprent, P. y Smeeton, N. C. (2007): *Applied Nonparametric Statistical Methods* (4ª edición). Chapman & Hall/CRC.
Silverman, B.W., (1986): *Density Estimation for Statistics and Data Analysis*, Chapman and Hall.

i) Inferencia Bayesiana

Aitchison, J. y Dunsmore, I. R. (1975): *Statistical Prediction Analysis.* Cambridge University Press.

Antelman, G. (1997): *Elementary Bayesian Statistics.* Edward Elgar Publishing.

Bernardo, J. M. y Smith, A. F. M. (2000): *Bayesian Theory*. John Wiley & Sons.

Berry, D. A. (1996): *Statistics: A Bayesian Perspective*. Duxbury Press.

Box, G. E. P. y Tiao, G. C. (1992): *Bayesian Inference in Statistical Analysis*. Wiley-Interscience.

Carlin, B. P. y Louis, T. A. (2000): *Bayes and Empirical Bayes methods for data analysis* Chapman and Hall/ CRC Press.

Congdon, P. (2007): *Bayesian Statistical Modelling* (2ª edición) **John Wiley & Sons.**

Gamerman, D. y Lopes, H. F. (2006): *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference* (2ª edición). Chapman & Hall/CRC.

Gelman, A., *et al.* (2003): *Bayesian Data Analysis* (2ª edición). Chapman & Hall/CRC.

*Hartigan, J. A. (1983): *Bayes theory*. Springer-Verlag.

Jeffreys, H. (1998): *Theory of Probability* (3ª edición). Oxford University Press, USA.

Lee, P. M. (2004): *Bayesian Statistics: An Introduction* (3ª edición). Hodder Arnold.

Lindley, D. V. (1970): *Introduction to Probability and Statistics from a Bayesian Viewpoint. Part 1: Probability; Part 2: Inference.* Cambridge University Press

— (1987): *Bayesian Statistics, a Review.* Society for Industrial Mathematics.

Mosteller, F. y Wallace, D. L. (1984): *Applied Bayesian and Classical Inference: The Case of the Federalist Papers* (2ª edición). Springer.

O'Hagan, A. y Forster, J. (2004): *Kendall's Advanced Theory of Statistics: Volume 2B: Bayesian Inference* (2ª edición). Hodder Arnold.

Phillips, L. D. (1973): *Bayesian Statistics for Social Scientists.* Thomas Nelson & Sons.

Press, S. J. (2002): *Subjective and Objective Bayesian Statistics: Principles, Models, and Applications* (2ª edición) . Wiley-Interscience.

Robert, C. P. (2007): *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation* (2ª edición). Springer Verlag.

Robert, C. P. y Casella, G. (2005): *Monte Carlo Statistical Methods* (2ª edición). Springer.

Savage, L. J. (1972): *The Foundations of Statistics* (2ª edición). Dover Publications.

Winkler, R. L. (2003): *An Introduction to Bayesian Inference and Decision* (2ª edición). Probabilistic Publishing.

Zellner, A. (1997): *Bayesian Analysis in Econometrics and Statistics*. Edward Elgar Publishing.

— (2007): *Information Processing and Bayesian Analysis.* Wiley-Interscience.

j) Decisión

Bell, D. E., Keeney, R. L., y Raiffa, H. (1977): *Conflicting Objectives in Decisions*. John Wiley and Sons.

Berger, J. O. (1993): *Statistical Decision Theory and Bayesian Analysis* (2ª edición). Springer.

*Ferguson, T. S. (1967): *Mathematical Statistics: A Decision Theoretic Approach.* Academic Press.

French, S. y Rios Insua, D. (2000): *Statistical Decision Theory*. Hodder Arnold.

Heinze, D. C. (1973): *Statistical decision analysis for management.* Grid.
Jedamus, P. (1978): *Statistical Analysis for Business Decisions*. McGraw-Hill.
Keeney, R. L. y Raiffa, H. (1993): *Decisions with Multiple Objectives: Preferences and Value Trade-Offs.* Cambridge University Press.
Lindgren, B. W. (1971): *Elements of Decision Theory.* Macmillan Company .
Lindley, D. V. (1991): *Making Decisions* (2ª edición). John Wiley & Sons.
Pratt, J. W., Raiffa, H., y Schlaifer, R. (2008): *Introduction to Statistical Decision Theory*. The MIT Press.
Raiffa, H. (1997): *Decision Analysis.* McGraw-Hill College.
Raiffa, H. y Schlaifer, R. (2000): *Applied Statistical Decision Theory.* Wiley-Interscience.
Ríos, S., Ríos Insua, S., y Ríos Insua, M. J. (1989): *Procesos de decisión multicriterios.* Ediciones de la Universidad Complutense de Madrid.
Schlaifer, R. (1969): *Analysis of Decisions Under Uncertainty.* McGraw-Hill.
— (1981): *Probability and statistics for business decisions: An introduction to managerial economics under uncertainty*. R.E. Krieger Pub.

k) Control de calidad

Besterfield, D. H. (2008): *Quality Control* (8ª edición). Prentice Hall.
Bowker, A. H., *et al.* (1981): *Estadística para ingenieros.* Prentice-Hall.
Braverman, J. D. (1981): *Fundamentals of Statistical Quality Control.* Prentice Hall.
Breyfogle, F. W. (2003): *Implementing Six Sigma: Smarter Solutions Using Statistical Methods* (2ª edición). John Wiley & Sons.
Charbonneau, H. C. y Webster, G. L. (1983): *Control de calidad.* Interamericana.
Deming, W. E. (2000): *Out of the Crisis.* The MIT Press.
Duncan, A. J. (1986): *Quality Control and Industrial Statistics* (5ª edición). Richard D. Irwin.
Grant, E. L. y Leavenworth, R. S. (1986): *Control estadístico de calidad.* Compañía Editorial Continental.
Hald, A. (1981): *Statistical Theory of Sampling Inspection by Attributes*. Academic Press.
Harry, M. P. y Schroeder, R. (2006): *Six Sigma: The Breakthrough Management Strategy Revolutionizing the World's Top Corporations.* Doubleday Business.
Ishikawa, K. (1990): *Guide to Quality Control.* Quality Resources.
Juran, J., *et al.* (1983): *Manual de control de la calidad.* Reverté.
Juran, J. M., Gryna, F. M., y Warleta, I. (1981): *Planificación y análisis de la calidad.* Reverté.
Ledolter, J. y Burrill, C. W. (1998): *Statistical Quality Control: Strategies and Tools for Continual Improvement.* John Wiley & Sons.
Montgomery, D. C. (2008): *Introduction to Statistical Quality Control* (6ª edición). John Wiley & Sons.
Otto, E. R. (1975): *Process Quality Control.* McGraw-Hill.
Prat, A., *et al.* (2004): *Métodos estadísticos: Control y mejora de la calidad.* Edicions UPC.
Pyzdek, T. (1999): *The Complete Guide to Six Sigma.* Quality Pub.
Ross, P. J. (1995): *Taguchi Techniques for Quality Engineering* (2ª edición). McGra-Hill.

Taguchi, G. (1981): *On-line quality control during production.* Japanese Standards Association.
Taguchi, G., Chowdhury, S., y Wu, Y. (2004): *Taguchi's Quality Engineering Handbook.* Wiley-Interscience.
Tiao, G. C., *et al.,* editores (2005): *Box on Quality and Discovery: with Design, Control, and Robustness.* Wiley-Interscience.
Wadsworth, H. M., Stephens, K. S., y Godfrey, A. B. (2001): *Modern Methods For Quality Control and Improvement* (2ª edición). John Wiley & Sons.
Wetherill (1977): *Sampling Inspection and Quality Control (Science Paperbacks)* (2ª edición). Chapman & Hall.

l) Libros de problemas y casos

Casas, J. M., García, C., y Zamora, A. I. (1998): *Problemas de Estadística: Descriptiva, probabilidad e inferencia.* Pirámide.
Chatterjee, S., Handcock, M. S. y Simonoff J.S. (1998): *A Casebook for a First Course in Statistics and Data Analysis* John Wiley & Sons.
Cuadras, C. M. (2000): *Problemas de probabilidades y estadística.* EUB.
García del Valle, T. y Martínez, A. (1996): *Problemas de Estadística.* Universidad del País Vasco.
Juan, J. (2002): *Problemas resueltos de Estadística.* Síntesis.
Kazmier, L. J. (1978): *Teoría y problemas de estadística aplicada a la administración y la economía.* McGraw-Hill.
Lipschutz, S. y Schiller, J. (1998): *Introduction to Probability and Statistics.* Schaum.
Montero, J. (1988): *Ejercicios y problemas de cálculo de probabilidades.* Díaz de Santos.
Romera, M. R. y Alonso, M. C. (1992): *Problemas de probabilidades y estadística.* Fundación General de la Universidad Politécnica de Madrid.
Ríos, S. (1989): *Ejercicios de Estadística.* Paraninfo.
Sarabia, A. y Maté, C. (1993): *Problemas de probabilidad y estadística.* Clag S.A.
Spiegel, M. R. y Stephens, L. J. (2004): *Estadistica.* McGraw-Hill.

Índice analítico